Lecture Notes in Computational Vision and Biomechanics 30

Durai Pandian
Xavier Fernando
Zubair Baig
Fuqian Shi *Editors*

Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)

Springer

Lecture Notes in Computational Vision and Biomechanics

Volume 30

The research related to the analysis of living structures (Biomechanics) has been a source of recent research in several distinct areas of science, for example, Mathematics, Mechanical Engineering, Physics, Informatics, Medicine and Sport. However, for its successful achievement, numerous research topics should be considered, such as image processing and analysis, geometric and numerical modelling, biomechanics, experimental analysis, mechanobiology and enhanced visualization, and their application to real cases must be developed and more investigation is needed. Additionally, enhanced hardware solutions and less invasive devices are demanded.

On the other hand, Image Analysis (Computational Vision) is used for the extraction of high level information from static images or dynamic image sequences. Examples of applications involving image analysis can be the study of motion of structures from image sequences, shape reconstruction from images, and medical diagnosis. As a multidisciplinary area, Computational Vision considers techniques and methods from other disciplines, such as Artificial Intelligence, Signal Processing, Mathematics, Physics and Informatics. Despite the many research projects in this area, more robust and efficient methods of Computational Imaging are still demanded in many application domains in Medicine, and their validation in real scenarios is matter of urgency.

These two important and predominant branches of Science are increasingly considered to be strongly connected and related. Hence, the main goal of the LNCV&B book series consists of the provision of a comprehensive forum for discussion on the current state-of-the-art in these fields by emphasizing their connection. The book series covers (but is not limited to):

- Applications of Computational Vision and Biomechanics
- Biometrics and Biomedical Pattern Analysis
- Cellular Imaging and Cellular Mechanics
- Clinical Biomechanics
- Computational Bioimaging and Visualization
- Computational Biology in Biomedical Imaging
- Development of Biomechanical Devices
- Device and Technique Development for Biomedical Imaging
- Digital Geometry Algorithms for Computational Vision and Visualization
- Experimental Biomechanics
- Gait & Posture Mechanics
- Multiscale Analysis in Biomechanics
- Neuromuscular Biomechanics
- Numerical Methods for Living Tissues
- Numerical Simulation
- Software Development on Computational Vision and Biomechanics
- Grid and High Performance Computing for Computational Vision and Biomechanics
- Image-based Geometric Modeling and Mesh Generation
- Image Processing and Analysis
- Image Processing and Visualization in Biofluids
- Image Understanding
- Material Models
- Mechanobiology
- Medical Image Analysis
- Molecular Mechanics
- Multi-Modal Image Systems
- Multiscale Biosensors in Biomedical Imaging
- Multiscale Devices and Biomems for Biomedical Imaging
- Musculoskeletal Biomechanics
- Sport Biomechanics
- Virtual Reality in Biomechanics
- Vision Systems

More information about this series at http://www.springer.com/series/8910

Durai Pandian · Xavier Fernando
Zubair Baig · Fuqian Shi
Editors

Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)

Springer

Editors
Durai Pandian
SCAD Institute of Technology
Palladam, India

Xavier Fernando
Department of Electrical
and Computer Engineering
Ryerson University
Toronto, ON, Canada

Zubair Baig
School of Computer and Security
Science
Edith Cowan University
Joondalup, WA, Australia

Fuqian Shi
Wenzhou Medical University
Wenzhou, China

ISSN 2212-9391 ISSN 2212-9413 (electronic)
Lecture Notes in Computational Vision and Biomechanics
ISBN 978-3-030-00664-8 ISBN 978-3-030-00665-5 (eBook)
https://doi.org/10.1007/978-3-030-00665-5

Library of Congress Control Number: 2018954619

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Contents

Digital Image Watermarking Using Sine Transform Technique

S. N. Prajwalasimha, S. Sonashree, C. J. Ashoka and P. Ashwini

Abstract In this chapter, digital image watermarking using sine transformation has been introduced in the frequency domain. In the proposed technique, the secret image is first compressed using a sine transformation and then embedded into the host image. The resultant images obtained from the proposed algorithm are subjected to various security attacks and the results are compared with other existing algorithms. The results obtained are better compared to the existing techniques.

1 Introduction

Authentication and information security play a vital role in multimedia communication. The digital watermarking technique provides both authentication and security to host data as well as a watermark, respectively. Many algorithms are designed to provide security to the information by encrypting the secret data. In order to provide authentication to the host, an information should be embedded into it [1–3]. Visible and invisible watermarking techniques are the fundamental classifications. While embedding the watermark into the host, the size of the watermark is a major aspect. Large the size of the watermark, more the vulnerability to the noise [4]. Irreversibility is the major problem among the conventional techniques, which leads to the large difference between the original and recovered watermark. This is unacceptable to the high authentication sectors like military and forensics. Reversible watermarking techniques provide the solution to the above-stated problem and are classified based on compression, differential expansion, and histogram shifting. In the compression-based algorithms, the secret image or watermark is first subjected to compression and then embedded into the host. Celik et al. proposed a lossless least significant bit (LSB)-based embedding algorithm [5]. In this technique, the compressed watermark is embedded into the LSB of the host. The correlation between the host and watermarked data considerably differs to some extent due to embedded information

S. N. Prajwalasimha (✉) · S. Sonashree · C. J. Ashoka · P. Ashwini
Department of Electronics & Communication, ATME College of Engineering, Mysuru, India
e-mail: prajwalasimha.sn1@gmail.com

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_1

in the host. A unique difference expansion based watermarking technique has been described by Tian [6]. In this approach, the embedded watermark in the host is very sensitive to noise. Histogram shifting method was proposed by Ni Z. et al., where the embedding process was done based on the peak values in the histograms of host and watermark [7]. Due to the histogram shifting, the pixel intensity variations can be observed in the watermarked data which intern affects the correlation between the host and watermarked data.

The proposed technique is a reversible, lossless and compression-based approach by which desired outcomes are noticed compared to popular existing techniques. Along with this, it supports secrete key authentication at the initial stage of the algorithm and provides the first stage security.

2 Proposed Scheme

The proposed algorithm consists of two phases: Embedding the watermarking into the host image and watermark extraction from the host image. Figure 1 shows the flow diagram of the proposed watermarking scheme.

2.1 Embedding the Watermark

Step 1 A Grayscale 256 × 256 image is taken as host image for the embedding process.

Step 2 By getting the desired secrete key, the watermark of size 128 × 128 is subjected for sine transformation.

$$W|(x, y) = \mathrm{DST}[W(x, y)] \quad 1 < x, y < 128 \tag{1}$$

where

$W(x, y)$ is the input watermark.
$W|(x, y)$ is the transformed watermark.

Step 3 A complete watermark of size 256 × 256 is composed by concatenating the transformed watermark back to back and side by side.

$$W||(x, y) = \begin{bmatrix} W|(x, y) & W|(x, y) \\ W|(x, y) & W|(x, y) \end{bmatrix} \tag{2}$$

where

$W|(x, y)$ is the transformed watermark of size 128 × 128.

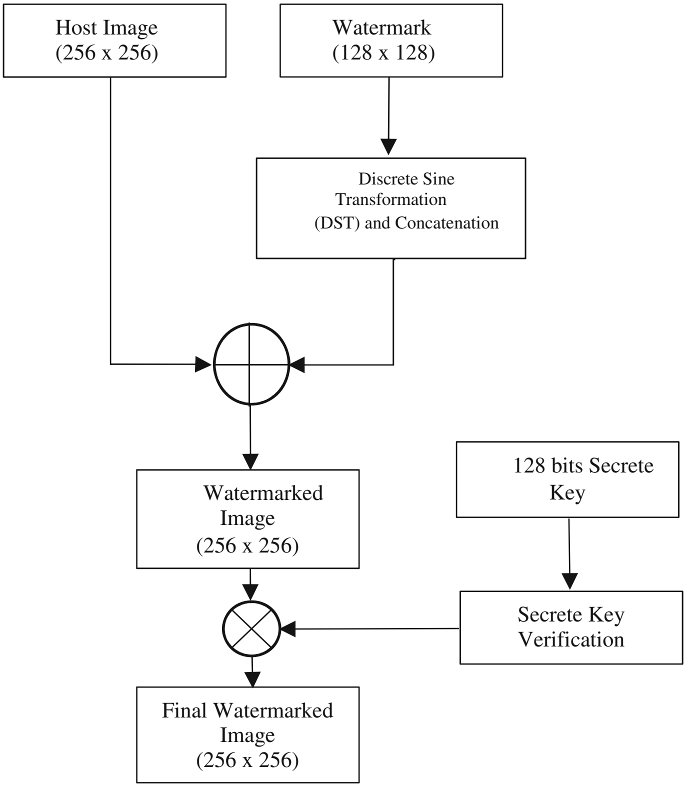


Fig. 1 Flow diagram of the proposed watermarking scheme

$W||(x, y)$ is the concatenated watermark of size 256×256.

Step 4 The arithmetic addition is performed between concatenated watermark and the host to get the final watermark.

$$W, (x, y) = H(x, y) + W||(x, y) \quad 1 < x, y < 256 \tag{3}$$

where

$W, (x, y)$ is the Watermarked image.
$H(x, y)$ is the host image of size.
$W||(x, y)$ is the transformed watermark.

2.2 De-watermarking

Step 1 Watermarked image is subjected to filtering in the first stage to reduce the noise contents.

Step 2 The filtered image is subjected to arithmetic subtraction with the host image to get the watermark in the transformed phase.

$$W'||(x, y) = W, (x, y) - H(x, y) \quad 1 < x, y < 256 \tag{4}$$

where

$W, (x, y)$ is the watermarked image.
$H(x, y)$ is the host image of size.
$W'||(x, y)$ is the transformed watermark.

Step 3 The obtained watermark of size 256 × 256 is separated by de-concatenating the transformed watermark.

$$W'||(x, y) = \begin{bmatrix} W'|(x, y) \; W'|(x, y) \\ W'|(x, y) \; W'|(x, y) \end{bmatrix} \tag{5}$$

where

$W'|(x, y)$ is the transformed watermark of size 128 × 128.
$W'||(x, y)$ is the concatenated watermark of size 256 × 256.

Step 4 By getting the desired secret key, the obtained watermark of size 128 × 128 in the transformed phase is subjected for inverse sine transformation.

$$W'|(x, y) = \text{IDST}[W||(x, y)] \quad 1 < x, y < 128 \tag{6}$$

where

$W||(x, y)$ is the transformed watermark.
$W'|(x, y)$ is inverse sine-transformed watermark.

Step 5 Comparison is made between the actual secret key and the input key and based on the desired combination, the final watermark is made visible.

3 Experimental Results

MATLAB2013a software is used for the analysis of the proposed algorithm. Three standard images are considered for the analysis with a secret image. The performance analysis involves Peak Signal-to-Noise Ratio (PSNR), Mean Square Error (MSE),

Table 1 The comparison of PSNR between the host and watermarked images

Images	PSNR (dB)
Lena	49.29 (DWT) [8]
	29.589 (DCT) [8]
	37.27 (HFT) [8]
	41.28 (IMD-WC-T) [8]
	40.6926 (Genetic algorithm) [9]
	51.6317 (Proposed algorithm)
Cameraman	40.2608 (Genetic algorithm) [9]
	51.6033 (Proposed algorithm)
Pirate	52.3612 (Genetic algorithm) [9]
	53.3748 (Proposed algorithm)

Table 2 The comparison of mean square error (MSE) between host and watermarked images

Images	MSE
Lena	1 (DWT) [8]
	26 (DCT) [8]
	12 (HFT) [8]
	5 (IMD-WC-T) [8]
	0.4466 (Proposed algorithm)
Cameraman	0.4495 (Proposed algorithm)
Pirate	0.2990 (Proposed algorithm)

Table 3 The comparison of correlation between the host and watermarked images

Images	Correlation
Lena	1 (DWT) [8]
	0.984 (DCT) [8]
	0.994 (HFT) [8]
	0.999 (IMD-WC-T) [8]
	1 (Proposed algorithm)
Cameraman	1 (Proposed algorithm)
Pirate	1 (Proposed algorithm)

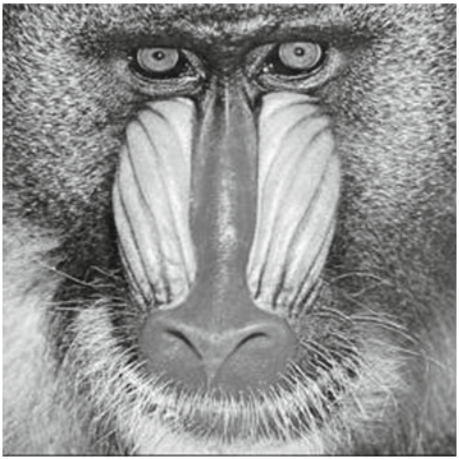
Correlation (NC), and Watermark to Document Ratio (WDR) between host image and watermarked image. In all the above tests, the proposed system has given much better results compared to existing algorithms as tabulated in (Tables 1, 2, 3 and 4; Figs. 2, 3, 4 and 5).

Table 4 The comparison of watermark to document ratio between host and watermarked (WDR) images

Images	WDR (dB)
Lena	−69.8801 (Proposed algorithm)
Cameraman	−69.8516 (Proposed algorithm)
Pirate	−71.6231 (Proposed algorithm)

Fig. 2 Watermark

Fig. 3 Host image



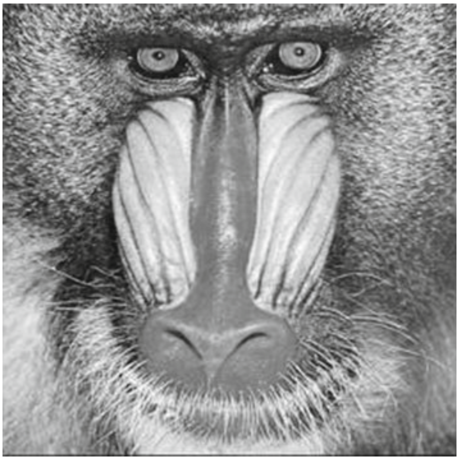


Fig. 4 Watermarked image

Fig. 5 Recovered watermark

4 Conclusion

In the proposed algorithm, the watermark is subjected for sine transformation and then embedded into the host image. The watermarked image is then subjected to various security tests. By this method, the embedded watermark is very less affected by robust attacks. It is observed that very high PSNR is achieved for the watermarked image compared to most popular existing techniques. Very less mean square error is observed and it very close to the ideal value. It is also noticed that the correlation between the host and watermarked images is equal to unity, which is observed only in the DWT technique. The statistical analysis indicates that the proposed algorithm is a better alternative to the existing techniques. And also the algorithm utilizes very less computational time for execution. Further, a hybrid technique can be developed by implanting both sine and cosine transformation to get better results.

References

1. Ahmaderaghi B, Kurugollu F, Del Rincon JM, Nekrasov D, Bouridane A (2018) Blind image watermark detection algorithm based on discrete Shearlet transform using statistical decision theory. IEEE Trans Comput Imaging 4(1):46–59
2. Brandão AS, Jorge DC (2016) Artificial neural networks applied to image steganography. IEEE Latin Am Trans 14(3):1361–1365
3. Prajwalasimha SN (2018) Pseudo-Hadamard Transformation-Based Image Encryption Scheme. Integrated Intelligent Computing, Communication and Security, information & communication technology, Studies in Computational Intelligence 771, Springer, pp 575–583
4. Prajwalasimha SN, Pavithra AC (2018) Digital Image Watermarking based on Successive Division. In: Proceedings of IEEE international conference on communication and electronics systems, pp 31–35
5. Celik MU, Sharma G, Tekalp AM, Saber E (2005) Lossless generalized-LSB data embedding. IEEE Trans Image Process 14(2):253–266
6. Tian J (2003) Reversible data embedding using a difference expansion. IEEE Trans Circuits Syst Video Technol 13(8):890–896
7. Ni Z, Shi YQ, Ansari N, Su W (2006) Reversible data hiding. IEEE Trans Circuits Syst Video Technol 16(3):354–362
8. Salama AS, Mokhtar MA (2016) Combined technique for improving digital image watermarking. In: Proceedings of IEEE international conference on computer and communications, pp 557–562
9. Khanna AK, Roy NR, Verma B (2017) Digital image watermarking and its optimization using genetic algorithm. In: Proceedings of IEEE international conference on computing, communication and automation, pp 1140–1144

Logarithmic Transform based Digital Watermarking Scheme

S. N. Prajwalasimha, A. N. Sowmyashree, B. Suraksha and H. P. Shashikumar

Abstract In this chapter, a new frequency domain transformation based digital watermarking technique has been introduced. The proposed technique adopts logarithmic transformation in order to embed the watermark image into the host image. The algorithm is subjected for various security attacks and compared with other frequently available digital watermarking schemes. The results obtained from the proposed algorithm are more satisfied compared to the existing systems.

1 Introduction

Data confidentiality and authentication are important aspects in the secured communication systems. Confidentiality can be more effectively achieved by various cryptosystems. In order to achieve both confidentiality and authentication, various steganosystems are available. As encryption, digital watermarking plays a vital role to achieve high degree of confidentiality and authentication [1–3].

Digital watermarking techniques are classified into two categories: spatial domain and frequency domain based on algorithms used for watermarking process [4, 5]. In spatial domain watermarking systems, the pixel values of the watermark image are directly altered and embedded into the host image. Least significant bit (LSB) technique is one among them. Whereas in frequency domain, the pixel values of the watermark image are first subjected for transformation and then altered before embedding into the host image. Based on applications, digital watermarking techniques are classified into two categories: robust and fragile techniques. Robust techniques are used for copyright and fingerprint protection and fragile techniques are used for data authentication and tamper detections [6]. Digital watermarking techniques are further classified into two classes: Perceptible and Imperceptible techniques based on visibility of the watermark on the host images. If the watermark is

S. N. Prajwalasimha (✉) · A. N. Sowmyashree · B. Suraksha · H. P. Shashikumar
Department of ECE, ATME College of Engineering, Mysuru, Karnataka, India
e-mail: prajwalasimha.sn1@gmail.com

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_2

visible on the host image, the technique used is termed as perceptible. Otherwise, it is imperceptible watermark [4, 7].

The proposed technique is a frequency domain, robust and imperceptible watermarking technique, which adopts logarithmic transformation to alter the pixel values of the watermark image and then it is embedded into the host image.

2 Related Work

Embedding a watermark into the multimedia data can be done by various methods. In the genetic algorithm based watermarking technique, strength of the algorithm is decided by three major factors: invisibility, less computations and resistance to various attacks. The algorithm is tested against filtering, resizing and Gaussian noise attacks. The results prove that the algorithm is better against the above mentioned attacks on various conditions [4]. More popular frequency domain watermarking algorithms are based on discrete cosine and wavelet transformation technique (DCT and DWT). A combined watermarking technique (IMD-WCT) has been proposed by Ahmed S. Salama et al. In this, they compared the results of various digital watermarking schemes with combined technique introduced by them. The results show that combined method is not much effective against discrete wavelet transformation technique [6].

Sanjay Kumar et al. made a comparative study on various watermarking schemes and classified them on their respective domains. In addition, they compared the results obtained by various algorithms by subjecting them to various attacks [7].

In interleaving and block extraction based technique, the watermarked image is subjected for various attacks including median filtering and noising. The recovered watermark is interrupted to some extent but resistive to them [8].

Mashruha Raquib Mitashe et al. proposed an adoptive technique, which is the combination of Discrete Wavelet Transform (DWT), XieBeni clustering (XFCM) and Particle Swarm Optimization (PSO). The watermark embedding has been made by DWT, XFCM is used to locate the positions to embed the watermark into the host and PSO is used for preprocessing the host image. The algorithm is effective against various attacks but the peak signal-to-noise ratio (PSNR) obtained for the watermarked image is less compared to other frequency domain watermark approaches [9]. In the Hilbert transformation based watermarking technique, the watermark is subjected for transformation and then embedded into the host. Various attacks are made on the watermarked image and observed better results in the recovery process [10].

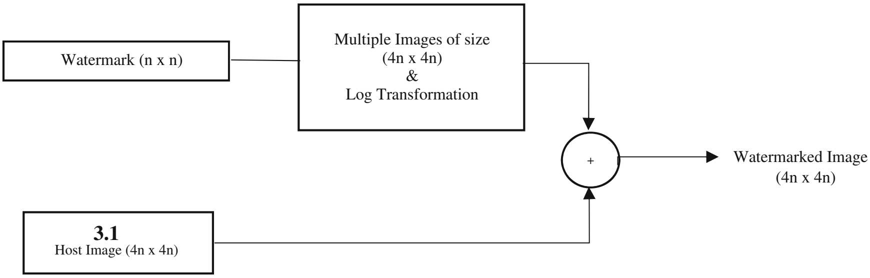


Fig. 1 Proposed model for watermark embeddeing algorithm

3 Proposed Scheme

The proposed algorithm is organized with both watermark embedding and extraction processes. A new method adopted here is logarithmic transformation. Each pixel value of the watermark is subjected for logarithmic transformation and the base is the ratio of maximum pixel value to its minimum. Figure 1 shows the proposed model of digital watermarking scheme.

3.1 Watermark Embedding Algorithm

Step 1 The input watermark image of size $n \times n$ as shown in Fig. 2 is made as multiple images of size $4n \times 4n$, which is same as the size of host image as shown in Fig. 3 ($4n \times 4n$).

Fig. 2 Watermark

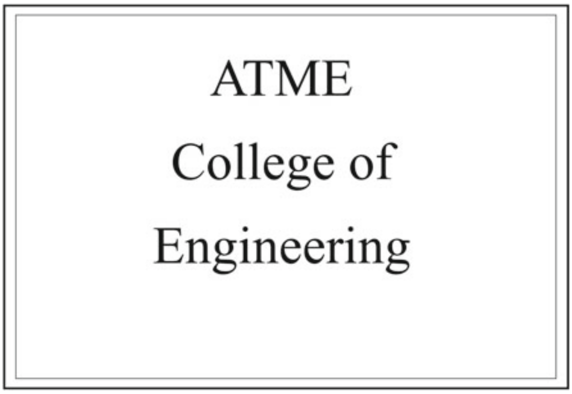


Fig. 3 Host image of Lena

Step 2 The watermark is then subjected to logarithmic transformation.

$$T(x, y) = \log_m[I(x, y)] \quad 0 < x, y < 4n \tag{1}$$

where $m = \frac{\max[I(x,y)]}{\min[I(x,y)]}$

$I(x, y)$ is the input watermark of size $4n \times 4n$
$T(x, y)$ is the transformed watermark of size $4n \times 4n$

Step 3 The transformed watermark and the host image are subjected to arithmetic addition.

$$W(x, y) = H(x, y) + T(x, y) \quad 0 < x, y < 4n \tag{2}$$

where $W(x, y)$ is the watermarked image of size $4n \times 4n$

$H(x, y)$ is the host image of size $4n \times 4n$
$T(x, y)$ is the transformed watermark of size $4n \times 4n$

The resultant watermarked image is of size $n \times n$ as shown in Fig. 4.

Fig. 4 Watermarked image

3.2 Watermark Extraction Algorithm

Step 1 The watermarked image is subjected for filtering in the first stage to reduce the noise contents.

Step 2 The filtered image is then subtracted with the host image to get the transformed watermark.

$$W'(x, y) = W^f(x, y) - H(x, y) \quad 0 < x, y < 4n \tag{3}$$

where $W'(x, y)$ is the same as $T(x, y)$ size $4n \times 4n$

$H(x, y)$ is the host image of size $4n \times 4n$
$W^f(x, y)$ is the transformed watermark after filtering of size $4n \times 4n$

Step 3 The transformed watermark is subjected to antilogarithmic transformation to get the original watermark.

$$I'(x, y) = m^{[W'(x,y)]} \quad 0 < x, y < 4n \tag{4}$$

where $m = \frac{\max[I(x,y)]}{\min[I(x,y)]}$

$I'(x, y)$ is the recovered watermark of size $4n \times 4n$

The resultant image is divided into eight $n \times n$ images and all together added to get the final retrieved watermark.

The recovered watermark is shown in Fig. 5.

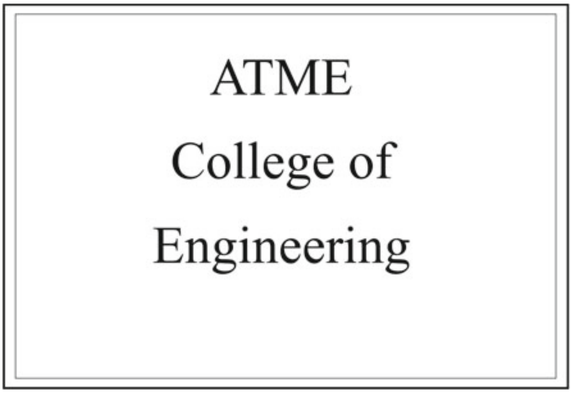


Fig. 5 Recovered watermark

4 Experimental Results

Three standard images of Lena, Cameraman, and Pirate are considered for the performance analysis [4]. The proposed algorithm is designed and implemented in MATLAB R2013a using Intel i3 processor @ 1.7 GHz, 4 GB DDR RAM and Windows 8 OS. The host images are considered with size 512 × 512 and the watermark of size 128 × 128. The performance analysis involves peak signal-to-noise ratio (PSNR), mean square error (MSE), Correlation (NC), and watermark to document ratio (WDR) between host image and watermarked image. In all the above tests, the proposed system has given much better results compared to existing algorithms as tabulated in Tables 1, 2, 3, 4 and 5.

Table 1 The comparison of PSNR between host and watermarked images

Images	PSNR (dB)
Lena	49.29 (DWT) [6]
	29.589 (DCT) [6]
	37.27 (HFT) [6]
	41.28 (IMD-WC-T) [6]
	40.6926 (Genetic algorithm) [4]
	54.0098 (Proposed algorithm)
Cameraman	40.2608 (Genetic algorithm) [4]
	54.0098 (Proposed algorithm)
Pirate	52.3612 (Genetic algorithm) [4]
	54.0098 (Proposed algorithm)

Table 2 The comparison of mean square error (MSE) between host and watermarked images

Images	MSE
Lena	1 (DWT) [6]
	26 (DCT) [6]
	12 (HFT) [6]
	5 (IMD-WC-T) [6]
	1.0372 (Proposed algorithm)
Cameraman	1.0372 (Proposed algorithm)
Pirate	1.0372 (Proposed algorithm)

Table 3 The comparison of correlation between the host and watermarked images

Images	Correlation
Lena	1 (DWT) [6]
	0.984 (DCT) [6]
	0.994 (HFT) [6]
	0.999 (IMD-WC-T) [6]
	1 (Proposed algorithm)
Cameraman	1 (Proposed algorithm)
Pirate	1 (Proposed algorithm)

Table 4 The comparison of watermark to document ratio between host and watermarked (WDR) images

Images	WDR (dB)
Lena	−68.9512 (Proposed algorithm)
Cameraman	−68.7148 (Proposed algorithm)
Pirate	−66.2388 (Proposed algorithm)

Table 5 Computational time of the proposed algorithm for standard images

Images	Computational time in seconds
Lena	0.308817
Cameraman	0.319627
Pirate	0.325353

5 Conclusion

In the proposed algorithm, the multiple images of the watermark are first transformed using logarithmic transformation and then embedded into the host image. By this method, the embedded watermark is very less affected by robust attacks. It is observed that very high PSNR is achieved for the watermarked image compared to most popular existing techniques. Very less mean square error is observed and it is almost equal to that of DWT technique. It is also noticed that the correlation between the host and watermarked images is equal to unity, which is observed in DWT technique. The statistical analysis indicates that the proposed algorithm is a better alternative to the exiting techniques. And also the algorithm utilizes very less computational time for execution. Further, the proposed technique can be combined with DWT to improve the results for the future research.

References

1. Nezhadarya E, Wang ZJ, Ward RK (2017) Robust image watermarking based on multiscale gradient direction quantization. IEEE Trans Inf Forensic Secur 6(4):1200–1213
2. Ahmaderaghi B, Kurugollu F, Del Rincon JM, Bouridane A (2018) Blind image watermark detection algorithm based on discrete shearlet transform using statistical decision theory. IEEE Trans Comput Imaging 4(1):46–59
3. Prajwalasimha SN (2018) Pseudo Hadamard transformation based image encryption scheme. In: Integrated intelligence computing, communication and security. Studies in computational intelligence, Springer, Chapter 58, Vol 771, pp 375–384
4. Khanna AK, Roy NR, Verma B (2017) Digital image watermarking and its optimization using genetic algorithm. In: Proceedings of international conference on computing, communication and automation (ICCCA2016), pp 1140–1144
5. Prajwalasimha SN, Shetter A, Swapna H (2018) Digital image watermarking using tenth root of exponential function. In: Proceedings of IEEE international conference on recent trends in electronics, information & communication technology, pp 634–637
6. Salama AS, Mokhtar MA (2016) Combined technique for improving digital image watermarking. In: Proceedings of 2nd IEEE international conference on computer and communications, pp 557–562
7. Kumar S, Dutta A (2016) Performance analysis of spatial domain digital watermarking techniques. In: Proceedings of international conference on information communication and embedded system (ICICES)
8. Bi H, Zhao C, Liu Y, Li N (2016) Digital watermarking based on interleaving extraction block compressed sensing in Contourlet domain. In: Proceedings of 9th international congress on image and signal processing, biomedical engineering and informatics (CISP-BMEI), pp 766–770
9. Mitashe MR, Habib ARB, Razzaque A, Tanima IA, Uddin J (2017) An adaptive digital image watermarking scheme with PSO, DWT and XFCM. In: Proceedings of IEEE international conference on imaging, vision & pattern recognition (ICIVPR)
10. Agarwal R, Santhanam MS, Srinivas K (2016) Digital watermarking: an approach based on Hilbert transform. In: Proceedings of international conference on computing, communication and automation (ICCCA), pp 1035–1042

A Survey of Medical Imaging, Storage and Transfer Techniques

R. R. Meenatchi Aparna and P. Shanmugavadivu

Abstract Medical data keeps growing with the growing number of scans every year. Patient experience plays a vital role in development of healthcare technologies. The speed with which the data can be accessed when the patient really wants to get diagnosed be it the same hospital or a different hospital becomes a very important requirement in future healthcare research. With growing amount of modality techniques and size of the captured images, it is very important to explore the latest technologies available to overcome bottlenecks. With (Computed Tomography) CT and Magnetic Resonance Imaging (MRI) modalities increasing the number of slices and size of the image captured per second, the diagnosis becomes accurate from the radiology perspective, but the need to optimize storage and transfer of the images without losing vital information becomes obviously evident. In addition security also plays an important role. There are various problems and risks when it comes to handling medical images because it is of key use to diagnose a disease which may be life threatening for the patient. There are evidences of radiologists waiting for the data for a considerable time to access the data for diagnosis. Hence time and quality plays a very important role in healthcare industry and it is major area of research which has to be explored. This scope of this survey is to discuss about the open issues and techniques to overcome the existing problems involved in medical imaging and transfer. This survey concludes the few optimization techniques with the medical imaging and transfer applications. Finally, limitation and future scope of improving medical imaging and transfer performance is discussed.

R. R. Meenatchi Aparna · P. Shanmugavadivu (✉)
Department of Computer Science and Applications, Gandhigram Rural Institute—Deemed University, Gandhigram, Dindigul, Tamilnadu, India
e-mail: psvadivu67@gmail.com

R. R. Meenatchi Aparna
e-mail: meenscience@gmail.com

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_3

1 Introduction

The aim of this study is to analyze the latest trends in medical imaging techniques and find how they can contribute to improve the medical image handling in terms of quality and performance. The processor speed, CPU (Central Processing Unit) utilization. Graphical Processing Unit (GPU) programming, Multiprocessing algorithms have proved to be a lot useful in other domains and has to be explored in full to improve the overall handling of images in a hospital. The following technologies are analyzed to find how they can contribute to improvement in performance and quality. Disadvantages of such techniques also open up areas of research which can become vital in future.

2 Medical Imaging Overview

In medical imaging data is highly important since it deals with patient privacy. This data is used for finding the patients disease and help him save his life. So data is important at that particular moment and at that particular stage of the disease. When this data is to be handled there are certain rules and regulations followed by the service providers, like Health Insurance Portability And Accountability Act (HIPAA).

There are standardization of the protocol for medical image exchange and storage. Most widely followed standard is Digital Imaging and Communication in Medicine (DICOM).

Yaorong et al. describe that image sharing by using CDs are a burden for the patients, and image sharing by networks increases the patient safety issues [1]. As explained in [2] it becomes inevitable to assure the privacy of the patient and integrity of the data when the data travels through different medium across geographical boundaries. This restriction is very important since we deal with medical data which becomes an integral part of patient's privacy information.

Hence it is very important to provide a high-quality image, within a specific time. With a highly accessible infrastructure for the hospital to cope up with the increase in medical data over the years and increased number of diseases which require scanning, currently the radiologists have to deal with terabytes of data unlike before. It adds to increasing number of cases [2]. Security of PHI is one of the topmost concerns of healthcare industry today. When patient data is processed, transported or archived, the security vulnerability loopholes play a major drive for the hackers to access the data.

There are certain representations where a non-DICOM image like SR reports can be stored, but the scope of this paper will be to address the problems dealt with DICOM images.

3 Healthcare Data

These are the data *volume* (size), typically in the petabyte range (1 PB = 10^{15} bytes), data *heterogeneity*, including (un)formatted, ASCII/Binary, (un)structured, and the data *velocity*, or data *derivative*, which captures the change, transfer, and discovery of raw and derived data [3–5].

4 Security of Healthcare Data

With the increasing trends of cloud computing there are some basic security and functional privacy issues come into picture especially with medical imaging workflows. The security of a medical imaging software is assured by allowing a particular user to access a particular patient data according to Health Insurance Portability and Accountability Act (HIPAA).

The major risks involved in medical images are

1. When the medical images are exposed to cloud, it is highly needed to adhere to security norms.
2. Required to anonymize the data.
3. Most of the cloud vendors according to current trend compress the data to achieve better efficiency but in a medical domain Compression is usually avoided to ensure data clarity
4. Eavesdropping the network is possible

Medical imaging requirements are highly risk related because of the following possibilities

- Complexity of Medical requirement
- Adherence to various international Standards
- Multimodality testing
- Risk involved in failure of the software (Affecting Human life)
- Loss of data cannot be entertained even to minimal extent
- Continuous evolution of the Technology with various sized and quantified images.

5 Whooping Increase in Healthcare Data

The volume, diversity, and velocity of biomedical data are exponentially increasing providing petabytes of new neuroimaging and genetics data every year [6]. Medical imaging solutions across the world generate humungous amount of data every second.

Based on the growing trend, it was estimated that over 100,000 terabytes of data will be performed in the United States during year 2014, which will generate petabytes of data [7]. When there is increase in the volume of data, medical image processing and analysis are computationally expensive [8].

6 DICOM

DICOM is the standard by which medical images are represented. The standard not only defines about the structure of the data but also about the messages used to exchange medical images. Each DICOM file has 128 bytes of header information and a set of tags t define a particular IoD (Information Object Definition). The IoDs are composed of a set of tags and values. The Tags are classified as mandatory type 1 tag or optional Type 2 tag.

There are various DIMSE (Dicom Messaging Service) service object pairs supported by every medical application which deals with message exchange mechanism of DICOM. The following are the major messages [9] used.

1. C ECHO
2. C MOVE
3. C STORE
4. C FIND

7 Latest Trends in Medical Imaging

The following technologies are mainly explored in correlation with handling medical images by researchers across the world:

1. Cloud Computing
2. Multicore
3. Big Data
4. Mobile Computing
5. GPU Programming

A. *Cloud computing*

Cloud computing addresses the problem of storing a large amount of data across different locations. Even though the same can be achieved by different techniques, the main attraction here is “pay as u use”. This can help the medical imaging vendors to reduce the cost to large extent.

At the same time the inherent advantage of cloud, a form of cloud in which the required hardware can be used based on necessity, i.e., "Infrastructure As A Service" will be of great help for the medical imaging vendors in terms of scalability and redundancy.

Although there are a lot of advantages by using cloud in medical imaging, there are certain problems when it comes to sharing medical images in server. As explained in [2], HIPAA(Health Insurance Portability and Accountability Act) plays a vital role in medical informatics. Every cloud vendor who supports the medical imaging applications should assure that the security rules are followed without any problem.

The security issues with cloud are

1. User access control for a particular data
2. Data reliability
3. Patient privacy issues
4. Data loss that might occur during anonymization and conversions.

Inspite of all these issues since the FDA (Food and Drug Administration) mandates the medical image data retention to 25 years [10], cloud will become the default solution for medical image archives [11]. Offsite backup, disaster recovery, high resource availability, and mobile device security and support are some of the greatest advantages that cloud can offer for medical imaging fraternity [12].

However, privacy and security concerns have slowed adoption of cloud storage, according to Nahim Daher, an analyst at consulting firm Frost & Sullivan. Cloud storage vendors store data at multiple sites, and the provider "doesn't know where the data is sitting and doesn't have direct oversight into who is looking at it," Daher says [13].

B. *Multicore*

The basic difference between an multiprocessor OS and a traditional OS being the management of process execution on different CPUs [14] creates new opportunities to do a lot of image processing operations at a faster pace which will again make the scans more faster and hence exposure to rays minimal.

With multicore technology being adapted to various industries, medical imaging can use multicore processing for process intensive tasks like image processing operations, 3D generation of models, simulation algorithms, etc. Sanjay et al. explains about how a multicore can be used effectively for image processing algorithms which will really speed up the algorithms [15].

There are some operations in which multicore becomes quite complex like sequential image viewing like a 2D or 3D movie, etc. Dev explains how a multicore Digital Signal processors (DSP s) inside the medical imaging equipment makes way for treatment to move towards patient instead of patient approaching the hospitals. This paves way for greater telemedicine opportunities. An application which is multi-threaded can be mapped to multicore, with each core performing an important part of the functionality.

In CT, Mike explains that the core count can go up to 1000, since there is a growing trend towards faster and more scan [16]. There are no proved evidences of multicore being used for image storage or transfer operations, which is still open for research.

C. *Big Data*

Big Data products like Cassandra really help the way the data is stored. With high and efficient data access, data redundancy also is addressed by some platforms where data is stored simultaneously at different locations and Data can be retrieved even when one location is down.

The main requirement when choosing a particular Big Data Technology is the updating of data should be done appropriately so that data consistency is maintained. With huge quantities of complex and high-quality data [17], medical data becomes obviously a candidate for Big Data.

The various algorithms used by the Big Data analytics pave way for diagnosis of the disease at a faster pace since it provides greater access to wider data. IN current time, Big Data analytics are explored hugely in US market for medical data [18] for further developments. Medical data include heterogeneous, multi-spectral, incomplete and imprecise observations (e.g., diagnosis, demographics, treatment, prevention of disease, illness, injury, and physical and mental impairments) derived from different sources using incongruent sampling [19].

There are a lot of companies starting to explore big data for medical images like explained in [20] mainly for the two reasons

1. Predictive analysis
2. Machine Learning

With huge amount of history medical data available, it would be possible in future to find out the disease that a patient might get in future, or make the machine to give data about the severity of a particular epidemic in a particular region. Big Data also helps to store unstructured data. This made medical image industry players to move towards Big data. Another reason was that medical data may consist of non-image data which is huge like waveforms, reports, etc.

D. *Mobile computing*

With latest mobile computing technology leading to smartphone support, medical images can also be shared through mobiles so that the radiologists can access the image “on the go” and help for investigation. With the tremendous development in telemedicine and teleradiology, the extent to which images are shared has increased over the last decade.

The main problems concerned with mobiles are in security of the patient images. All security issues related to mobile networks hold good for the medical images in Mobile computing technology also and hence can lead to access control issues.

E. *GPU programming*

This technology has seen a tremendous growth in last few years and research is extensively done to improve the usage of GPUs in programming. The main idea of this technology is that the GPU (graphical processing unit) which sits idle during most of the time can be utilized for intensive operations since there are more number of cores in a GPU. It is more powerful when compared to CPU and quite a set of constructs are available to implement GPU programming like CUDA, OpenMP 4.0, etc. Anders et al. describe that the denoising is performed in 8 min in GPU on contrary to CPU architectures where the performance was 50 min [21]. In a medical imaging application whenever the graphical operations are not done, the GPU can be effectively used for workflows like handling the data especially in image transfer workflows. The applications which use Geometric transformations and high level of mathematical transformations can very well use the GPU which is very much capable of doing multiprocessing operations [22].

Jeyarajan et al. describe about the mapping of a motion estimation algorithm to GPU architectures and explain the performance gains [23]. With developing languages such as CUDA it is possible to achieve high performance especially when the operations can be made parallel.

F. *Vendor Neutral Archives*

Vendor Neutral Archive represents the storage of medical data in a non proprietary format and provides for seamlessly sharing across organizations.

The main functionality of VNA is to decouple the vendor specific PACS from the DICOM network and replace it with a neutral module which can easily convert the images stored in Archive to DICOM and process further. By this means dependency on PACS and native format is completely prevented. PACS from one vendor can be easily replaced with another vendor.

8 Existing Works on Latest Technologies

Technology	Author, Year and References	Problem addressed
Cloud	Kagadis et al. [24]	Applicability of advanced cloud computing in medical imaging
	Karthikeyan et al. [25]	Emergency healthcare sector in an umbrella with physical secured patient records with SaaS
	Dai et al. [26]	Review extant cloud-based services in bioinformatics, classify
	Ya et al. [27]	Cloud-based Hospital Information Service Center
	Liu et al. [28]	Three-layer hybrid cloud
MultiCore	Niendorf et al. [29]	Cardiovascular MRI slowdown
	Lecron et al. [30]	Vertebra Detection and Segmentation in X-Ray Images, slowdown
	Xu et al. [31]	Iterative reconstruction algorithm
	Hofmann et al. [32]	FDK algorithm
	Mittal et al. [33]	CPU-GPU Heterogeneous Computing
GPU	Howison [34]	GPU ray casting
	Massanes et al. [35]	Multi-GPU proceeding
	Olmedo et al. [36]	Image processing
	Westhoff [37]	Segmentation
	Weinlich et al. [38]	Raycasting
VNA	Kumar et al. [39]	PACS neutrality
	Cook [40]	VNA
	Gray [41]	VNA

9 Proposed Work

GPU programming is already proved to be more efficient in terms of graphical processing. Our experiments would analyze the use of GPU in creation of DICOM messages post-image processing so that the performance can be improved. The main idea being using GPU is about 60% of the time in transfer is used for message preparation. So ideally if this part is made to be executed in GPU, there would be a heavy performance improvement.

The modules used in the transfer would be message creation module, sending module on the sending end and the receiving module and demultiplexing module on the receiving end in a network. Figure 1 shows how this would happen in a network.

Our proposal would be to move the message creation module to GPU so that the sending module becomes very balanced to use the network efficiently.

Figure 2 illustrates the proposed workflow.

The proposed workflow will create the messages in GPU and transfer it using the CPU. The next step of this experiment would be to use the multicore in the CPU as well by the sending module (Fig. 3).

Fig. 1 Normal workflow

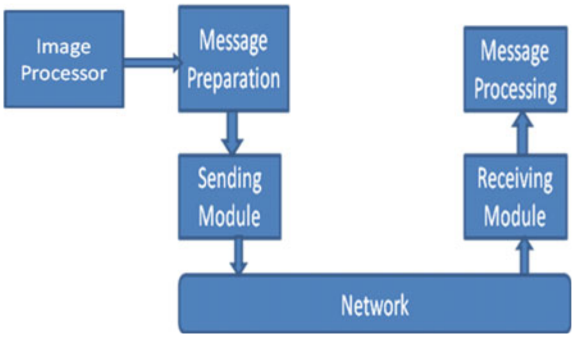
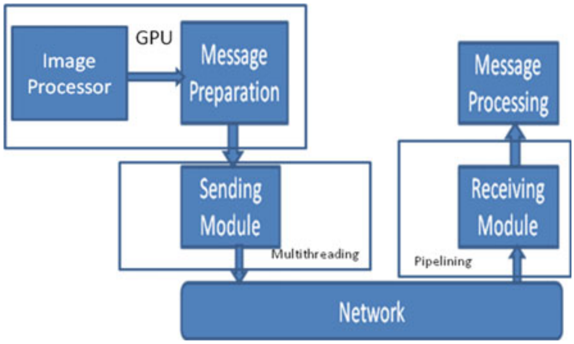


Fig. 2 Proposed workflow



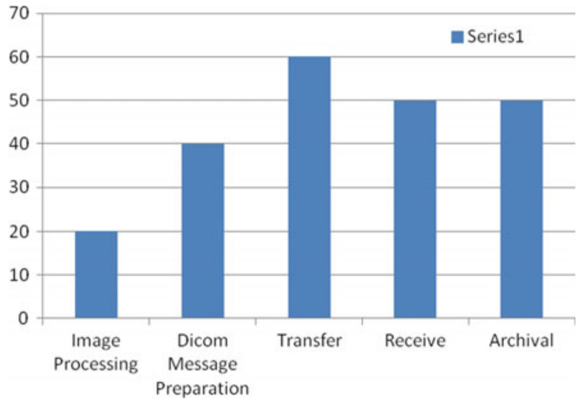


Fig. 3 Stages and the percentage of time spent in transfer

A. *Hardware Configuration*

The specification of the hardware used.

GPU:	NVIDIA GeForce 920 M
Speed:	1.4 GHz
Multiprocessor count:	2
Number of cores:	385 (2×192)
Memory:	4 GB
Threads:	1024 per Block
Language:	CUDA 7.0, OpenMP

B. *Tools Used*

The following are the list of tools used for this experiment:

1. OpenMP
2. C
3. C++
4. NVidia GPU
5. Pentium processor
6. JDICOM

10 Results

Since the GPU which is idle for most of the time is used for the CPU intensive operation, it leads to a better performance in the transfer workflow. The heterogeneous combination of multicore CPU and GPU implementation gives higher performance when compared to conventional transfer method. This overcomes the disadvantages of multithreading in which the program runs in the same processor in chunks by using time slicing and GPU programming, the memory swap that happens between GPU and CPU to communicate the output.

11 Conclusion

The major advantage of the above workflow is that we are using the most powerful processing units for image processing and message preparation and CPU for network send. A combination of hybrid strategy would provide a multi stage performance improvement.

With advancements in various medical imaging technologies, the area of medical imaging transfers has to be explored adapting to the new trends. Increasing the performance of medical transfers is a challenge which needs to be explored. This research will confirm performance improvement achieved by using heterogenous model for medical imaging and transfer.

References

1. Ge Y, Ahn DK, Unde B, Gage H, Carr JJ (2013) Patient-controlled sharing of medical imaging data across unaffiliated healthcare organizations. J Am Med Inf Assoc 20(1):157–163.http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3555338/
2. Medical imaging analytics (2015) (Online) Available https://www.research.ibm.com/haifa/dept/imt/mia.shtml. Accessed 17 Aug 2015
3. Foster K, Spicer M, Nathan S (2011) IBM infosphere streams: assembling continuous insight in the information revolution. International Technical Support Organization, San Jose, CA
4. Howe D et al (2008) Big data: the future of biocuration. Nature 455(7209):47–50
5. Lynch C (2008) Big data: how do your data grow? Nature 455(7209):28–29
6. Dinov ID, Petrosyan P, Liu Z, Eggert P, Zamanyan A, Torri F, Macciardi F, Hobel S, Moon SW, Sung YH, Toga AW (2014) The perfect neuroimaging-genetics-computation storm: collision of petabytes of data, millions of hardware devices and thousands of software tools. Brain Imaging Behav 8(2):311–322
7. Prepare for disaster & tackle terabytes when evaluating medical image archiving (2008) Frost & Sullivan. http://www.frost.com
8. Rodger JA (2015) Discovery of medical big data analytics: improving the prediction of traumatic brain injury survival rates by data mining patient informatics processing software hybrid hadoop hive. Inf Med Unlocked 1:17–26
9. DICOM standard 2015, Message Exchange, NEMA
10. Retention and storage of images and radiological patient data. Dated February 2008. https://www.rcr.ac.uk/docs/radiology/pdf/ITguidance_Retention_storage_images.pdf
11. Liu BJ, Cao F, Zhou MZ, Mogel G, Documet L (2003) Trends in PACS image storage and archive. Comput Med Imaging Graph 27

12. Shah D, Kollaikal P Top trends in medical imaging technology (Online) Available: http://www.citiustech.com/uploads/article/pdf/top-trends-in-medical-imaging-technology-89.pdf. Accessed 22 Jan 2018
13. Healthcare in cloud: a Storage solution or security risk. http://www.advisory.com/daily-briefing/2013/04/10/health-care-in-the-cloud-a-storage-solution-or-security-risk. Date 10 Apr 2013
14. Kagadis GC, Langer SG (2012) Informatics in medical imaging. CRC Press, Boca Raton
15. Saxena S, Sharma N, Sharma S (2013) Image processing tasks using parallel computing in multi core architecture and its applications in medical imaging. Int J Adv Res Comput Commun Eng 2(4)
16. Hinds M (2009) White paper on "Power up: moving toward parallel processing in medical imaging compute systems
17. Cowan B (2015) Big data medical imaging (Online) Available http://nihi.auckland.ac.nz/sites/nihi.auckland.ac.nz/files/pdf/informatics/bigdata/Big%20Data%20Medical%20Imaging%20-%20Brett%20Cowan%206.pdf. Accessed 17 Aug 2015
18. PRNewsWire (2015) US medical imaging industry leaps firmly into the big data realm (Online) Available http://www.prnewswire.com/news-releases/us-medical-imaging-industry-leaps-firmly-into-the-big-data-realm-300105491.html. Accessed 17 Aug 2015
19. Dinov ID (2016) Volume and value of big healthcare data. J Med Stat Inform 4:3. https://doi.org/10.7243/2053-7662-4-3
20. Ridley EL (2015) http://www.auntminnie.com. Israeli start-up eyes big-data tools for imaging analysis (Online) Available https://mail.google.com/mail/u/0/#inbox/14f3c7be7d177547?projector=1. Accessed 17 Aug 2015
21. Eklund A, Andersson M, Knutsson H (2011) True 4D image denoising on the GPU. Int J Biomed Imaging 2011
22. Shams R, Sadeghi P, Kennedy RA, Hartley RI (2010) A survey of medical image registration on multicore and the GPU. IEEE Sign Process Mag 27
23. Thiyagalingam J, Goodman D, Schnabel JA, Trefethen A, Grau V (2011) On the usage of GPUs for efficient motion estimation in medical image sequences. Int J Biomed Imag 2011
24. Kagadis GC, Kloukinas C, Moore K, Philbin J, Papadimitroulas P, Alexakos C, Nagy PG, Visvikis D, Hendee WR (2013) Cloud computing in medical imaging. Med Phys 40(7):070901. https://doi.org/10.1118/1.4811272
25. Karthikeyan N, Sukanesh R (2012) Cloud based emergency health care information service in India. J Med Syst 36(6):4031–4036. https://doi.org/10.1007/s10916-012-9875-6
26. Dai L, Gao X, Guo Y, Xiao J, Zhang Z (2012) Bioinformatics clouds for big data manipulation. Biol Direct 28(7):43. https://doi.org/10.1186/1745-6150-7-43 discussion 43
27. Yao Q, Han X, Ma XK, Xue YF, Chen YJ, Li JS (2014) Cloud-based hospital information system as a service for grassroots healthcare institutions. J Med Syst 38(9):104. https://doi.org/10.1007/s10916-014-0104-3
28. Liu L, Chen W, Nie M, Zhang F, Wang Y, He A, Wang X, Yan G (2016) iMAGE cloud: medical image processing as a service for regional healthcare in a hybrid cloud environment. Environ Health Prev Med 21(6):563–571
29. Niendorf T, Sodickson DK (2006) Parallel imaging in cardiovascular MRI: methods and applications. NMR Biomed 19(3):325–341
30. Lecron F, Mahmoudi SA, Benjelloun M, Mahmoudi S, Manneback P (2011) Heterogeneous computing for vertebra detection and segmentation in X-ray images. Int J Biomed Imaging 2011, Article ID 640208
31. Xu M, Thulasiraman P (2011) Mapping iterative medical imaging algorithm on cell accelerator. Int J Biomed Imaging 2011, Article ID 843924

32. Hofmann J, Treibig J, Hager G, Wellein G (2013) Performance engineering for a medical imaging application on the intel Xeon Phi accelerator (online) https://arxiv.org/pdf/1401.3615.pdf. Accessed 17 Dec 2013
33. Mittal S, Vetter JS (2015) A survey of CPU-GPU heterogeneous computing techniques. ACM Comput Surv (CSUR), 47(4), Article No. 69
34. Howison M (2010) Comparing GPU implementations of bilateral and anisotropic diffusion filters for 3D biomedical datasets. In: SIAM conferences of imaging science
35. Massanes F, Cadennes M, Brankov JG (2011) Compute-unified device architecture implementation of a block-matching algorithm for multiple graphical processing unit cards. J Electron Imaging 20(3):1–10
36. Olmedo E, Calleja J, Benitez A, Medina MA (2012) Point to point processing of digital images using parallel computing. IJCSI Int J Comput Sci Issues 9(3):1–10
37. Westhoff AM (2014) Hybrid parallelization of a seeded region growing segmentation of brain images for a GPU cluster. In: Proceedings of the international conferences on architecture of computing systems
38. Weinlich A, Keck B, Scherl H, Kowarschik M, Hornegger J (2008) Comparison of highspeed ray casting on GPU using CUDA and OpenGL. In: Proceedings of the international workshop on new frontiers in high-performance & hardware-aware computing, pp 25–30
39. Tapesh Kumar Agarwal, Sanjeev (2012) Vendor neutral archive in PACS. Indian J Radiol Imaging 22(4):242–245
40. Cook R Is VNA the future of image delivery? (online) http://www.healthcareitnews.com/news/should-you-use-vna-whats-vna
41. Gray M The bridge from PACS to VNA scale out (online) https://www.emc.com/collateral/hardware/white-papers/h10699-bridge-from-pacs-to-vna-wp.pdf

Majority Voting Algorithm for Diagnosing of Imbalanced Malaria Disease

T. Sajana and M. R. Narasingarao

Abstract Vector borne diseases like malaria fever is one of the most elevating issues in medical domain. Accurate identification of a patient from the given set of samples and classification becomes one of the challenging task when dealing with imbalanced datasets. Many conventional machine learning and data mining algorithms are shows poor performance to classify skewed distributed data because they are trained very well with the majority class samples only. Proposing an ensemble method called majority voting defined with a set of machine learning algorithms namely decision tree—C4.5, Naive Bayesian and K-Nearest Neighbor (KNN) classifiers. Classification of samples can be done based on the majority voting of classifiers. Experiment results stating that voting ensemble method shows classification accuracy of 95.2% on imbalanced malaria disease data whereas dealing with balanced malaria disease data voting ensembler shows 92.1% of accuracy. Consequently voting shows 100% classification report on precision, Recall and F1-Score on imbalanced malaria disease data sets whereas on balanced malaria disease data voting shows 96% of Precision, Recall and F1-Score metrics.

1 Introduction

Mosquito bite diseases like Malaria is one of the frightful diseases facing by many people throughout the world wide. Due to the climate changes and food habits still the disease has good influence on remote areas [1, 2]. Regarding health reports of various organizations the effect of vector borne diseases still become a challenging issue in medical domain [3] and Identification of an effected patient within time is very important among the set of patients data [3, 4]. Even in other applications like

T. Sajana (✉) · M. R. Narasingarao
Department of Computer Science & Engineering, K L E F, Vaddeswaram, Guntur, India
e-mail: sajana.cse@kluniversity.in

M. R. Narasingarao
e-mail: Ramanarasingarao@kluniversity.in

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_4

identification of oil spills in satellite images, spam emails detections are also facing the problem of classification of minority class samples [5, 6].

Malaria—caused by the bite of female mosquito called "Anopheles" which affects the red blood cells directly that causes the damage of liver functionality. There are mainly four types of malaria parasites namely *Plasmodium falciparum*, *Plasmodium vivax*, Plasmodium Malaria and *Plasmodium ovale* and additionally to these parasites *Plasmodium knowlesi* also to be considered for the identification of malaria disease [7, 8, 17].

Many authors are proposed various data mining and machine learning algorithms for the classification of malaria disease. But, all these traditional algorithms are well performed when there is balanced class distribution present in between the samples. But, in the area of imbalanced class problem all these algorithms are biasing towards majority class samples only which becomes a crucial task in identification of effected patient called minority class sample. Hence, consider ensemble methods which are designed with a set of classifiers and trained on given samples very effectively that inference the classifiers for the exact classification of effected patients [9, 10, 19–23].

The rest of the paper stated that history on imbalanced malaria disease and balanced malaria disease datasets which was described in Sect. 2. Section 3 explores the methodology for classification of both malaria disease data using voting ensemble method and experimental outcomes are discussed in Sect. 4.

2 Literature Review

Identification of an effected patient within time and diagnosing plays a vital role. Many authors are suggested, proposed conventional machine learning on equal distribution of class samples of malaria disease data as stated below:

- Pandit and Anand [11] proposed Artificial Neural Networks for the identification of effected erythrocytes.
- Bbosa et al. [12] suggested rule based classification for identification of effected malaria patient.
- Wu and Wong [13] proposed a neural network model effective diagnosing of malaria disease.
- Tsai et al. [14] suggested images segmentation method for detection of malaria parasites.
- Rahmanti et al. [15] investigated on parasite classification using image processing and KNN techniques.
- Charpe and Bairagi [16] identified the parasites stage using image processing and classification techniques.
- Somasekar and Reddy [17] derived a method for identification of effected erythrocytes based on Adaptive median filter, edge enhancement and Fuzzy C-Means clustering techniques.

But he traditional machine learning algorithms are biased towards the majority class samples only. Hence, due to the nature of imbalanced distribution, classification of minority class samples becomes a burning issue suggested by various authors [9, 10, 18–23].

Researchers are proposed many methods to handle the class imbalance problem im many applications including malaria disease. Consider the history of class imbalance problem in different applications:

- Guo Haixiang et al. presented a plethora of methods for handling imbalanced data sets in the era of many application domains [24].
- Salma Jamal developed a Predictive model of anti-malarial molecules inhibiting apicoplast formation for an imbalanced malaria disease dataset using Cost sensitive Naive Bayesian, Random Forest and J48 algorithms and found that Random Forest algorithm produces better accuracy [25].
- Bruno B. Andrade et al. suggested that severe stage of malaria disease causes to reducing of inflammatory cytokines which is a high level imbalance class problem in medical domain [26].
- Rashmi Dubey proposed an ensemble frame work for classification of Alzheimer's disease [27].
- Wing W. Y. Ng et al. Suggested Diversified Sensitivity-Based Under sampling method for handling different imbalanced datasets [28].
- Yazan F. et al. Experimentally investigated 2080 cardiac patients data using various classifiers for better disease diagnosing [29].
- Jia Pengfei et al. Developed a new sampling method called Distinct—Borderline algorithm for balancing the class distribution of imbalanced data sets [30].
- N. Poolsawad et al. Conducted a comparative study between set of classifiers like Multilayer Perceptron (MLP), RBFN, SVM, Decision Tree and Random forest algorithms on LIFELAB dataset and found that RF algorithm produces better accuracy [22].
- V. Garcia et al. İnestigated various methods for imbalanced data sets handling [31].
- Jaree Thongkam et al. proposed C-Support vector machine method for prediction of imbalanced breast cancer data set [32].
- Xing-Ming Zhao et al. Proposed an ensemble classifiers for protein classification. Conducted a two ways of comparative study between balanced data and unbalanced data using individual classifiers & ensemble classifiers and concluded that ensemble classifiers are shows best classification performance [33].
- Victoria López et al. proposed an ensemble framework on different datasets by conducting a comparison study between C4.5, SVM and instance based learning methods [34].
- Li Ma et al. investigated that CURE-SMOTE (Clustering Using Representatives—Synthetic Minority Oversampling Technique) is best sampling method over the Border line SMOTE1, safe level SMOTE, C-SMOTE, k-means SMOTE techniques and then suggested Random Forest, a hybrid algorithm for feature selection and parameter optimization [35].

3 Proposed Methodology

In the era of class imbalance problem many different methods are defined for classification of samples especially samples belongs to the minority class samples. But, the difficulty with these methods like data sampling is they can over fit the data or loss some valuable sample also which tends to misclassify the affected patient when compared with un affected patient. Hence, Ensemble methods are best methods for classification of imbalanced datasets because set of classifiers are grouped together to increase the classifiers accuracy performance and also they can reduce data over fitting which is one of the drawback of data handling techniques for imbalanced data. Another advantage is error rate will be automatically reduced or simply null. Consider the proposed methodology design as shown in Fig. 1 which describes voting ensembler a combination of Decision tree—C4.5, Naive Bayesian and KNN classifiers as stated below.

Consider the Decision tree - C4.5 algorithm:
Input: data set D.
Output: A Decision tree.
Method:

- Construct a node N if instances in D are belongs to same class, C, then Return N as a leaf node labelled with the class C;
- If attribute _ list is empty then
- Return N as a leaf node labelled with the majority class in D;
  // majority voting
- Apply attribute_selection_method (D, attribute _ list) to find the "best" splitting.
- Let node 'N' hold the tuples of partition D.
- Calculate information gain to classify a tuple in D is given by

$$\mathrm{Info(D)} = -\mathrm{i}^{*} \log 2\left(\mathrm{p_i}\right)$$

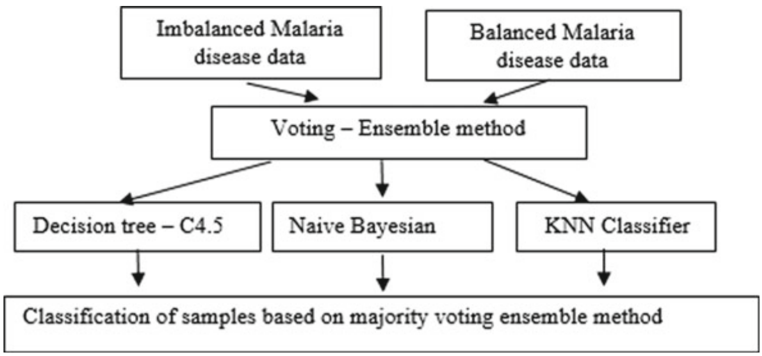


Fig. 1 Methodolgy for classification of samples using voting ensembler

- Select node N with maximum information gain.
- Find the expected information required to classify a tuple D based on partitioning of an attribute set A { a_1, a_2, a_3, a_4, …, a_v} is measured by

$$Info_A(D) = \sum_{j=1}^{v} \frac{|D_j|}{|D|} \times Info(D_j)$$

- Branch the sub trees for j
- let Dj be the set of data tuples in D satisfying outcome j;
- Amount of information attained by A is defined by

$$\mathrm{Gain(A)} = \mathrm{Info(D)} - \mathrm{Info_A(D)}$$

- Apply normalization method called split information to overcome the bias problem with information gain by an attribute A

$$SplitInfo_A(D) = -\sum_{j=1}^{v} \frac{|D_j|}{|D|} \times \log_2\left(\frac{|D_j|}{|D|}\right)$$

- Select the attribute with highest gain ratio,

$$GainRatio(A) = \frac{Gain(A)}{SplitInfo(A)}$$

- if Dj is NULL
- Then branch a node with majority class in D.
- Else grow the sub tree by creating a node which is selected by decision tree.
- Apply Pessimistic pruning method for tree pruning to find the error rate using training data set.

Consider the Naive Bayesian classification algorithm:
Input: dataset D.
Output: Classification of sample.
Method:

- D: A dataset with training tuples.
- $X = (x_1, x_2, x_3, \ldots, x_m)$ is an n-dimensional vector each vector tuple of D defined over 'n' attributes of A_1, A_2, A_3, … A_n.
- C_1, C_2 … Cm: set of classes.
- Find the maximum posterior probability of each sample belongs to class 'C_i' if and only if, $P(C_i|X) > P(C_j|X)$ for $1 \le j \le m$, j! = i.
- By Bayesian theorem calculate maximized posterior hypothesis,
- $P(C_i|X) = [P(X|C_i).\ P(C_i)] / P(X)$ where P(X) is constant.
- Predict the class of X based on class C_i if and only if $P(X|C_i)\ P(C_i) > P(X|C_j)\ P(C_j)$ for $1 \le j \le m, j \ne i$.

Consider the KNN Classifier algorithm as follows:
Input: dataset D.
Output: Classification of given data.
Method:

- Find the distance between points x and x_i i.e., $d(x, x_i)$ where $i = 1, 2, \ldots n$.
- Select initial 'm' distances from the set of increasing order of distances.
- Obtain k points w.r.t the $m = k$ distances.
- For $k \geq 0$, k_i defines the no of points belongs to i^{th} class.
- For each $k_i > k_j \; \forall \; i \neq j$ assign 'x' in class'i'.

Dataset : Gathered 165 tuples data (real dataset) defined as follows:

Attributes	Actual range
Age	1–20 years
Haemoglobin	M13.0–18.0/F11.0–16.5 g%
RBC	3.80–5.80 millions/cumm
Hct	35.0–50.0%
Mcv	80–97 fl
Mch	26.5–33.5 pg/cells
Mchc	31.5–35.0%
Platelets	1.50–3.90 lakhs/cumm
WBC	3500–10,000 cells/cumm
Granuls	43.0–76.0%
Lymphocytes	17.0–48.0%
Monocytes	4.0–10.0%
Malaria	Yes/No

Consider the samples distribution of collected Malaria Dataset as described in Table 1.

Table 1 Samples distribution and imbalanced % of malaria disease dataset

No. of tuples	Parameters	Minority/effected samples	Majority/uneffected Class samples	Class (effected/uneffected)	% Class (effected, unef-fected)
165	13	5	160	(Positive, Negative)	(3.1, 96.9)

4 Results and Discussions

Proposing a well-defined algorithm of ensemble method namely voting a combination of set of classifiers Decision tree—C4.5, Naive Bayesian and KNN which are trained with common set of samples to the frame work as shown in Fig. 1 and then performance of these algorithm measured with various metrics like accuracy, sensitivity, specificity and F1-score which are derived from confusion matrix as defined Table 2.

Presenting a combination of machine learning algorithms like Decision Tree—C4.5, Naive Bayesian and KNN Classifiers as an ensemble method of Voting which classifies the samples based on majority voting. Voting ensemble algorithm increases the classification accuracy performance of imbalanced malaria disease data i.e., 95.2% when compared with balanced malaria disease data as 92.1% which is as shown in Fig. 2. But unlike other ensemble methods voting of C4.5, Naive Bayesian, KNN Classifiers shows good bias variance at each k-fold cross validation of both imbalanced and balanced malaria disease datasets which is shown in Fig. 3 and also shows 100% performance on skewed malaria disease data w.r.t Precision, Recall and F1-Score as shown in Figs. 4, 5 and 6.

Table 2 Confusion Matrix

		Predicted	
		Effected/minority class	Uneffected/majority class
Actual	Effected class	TP	FN
	Uneffected class	FP	TN

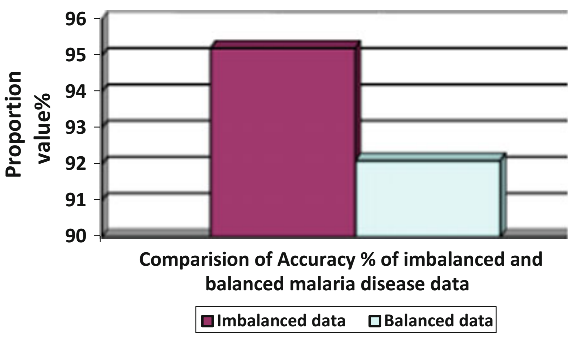


Fig. 2 Classification accuracy % of both imbalanced and balanced malaria disease data using voting ensemble algorithm

Fig. 3 Cross validation results of voting ensemble algorithm on both imbalanced and balanced malaria disease datasets

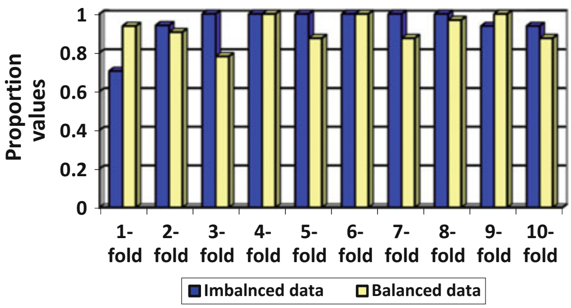


Fig. 4 Precision % of voting ensemble algorithm on both imbalanced and balanced malaria disease datasets

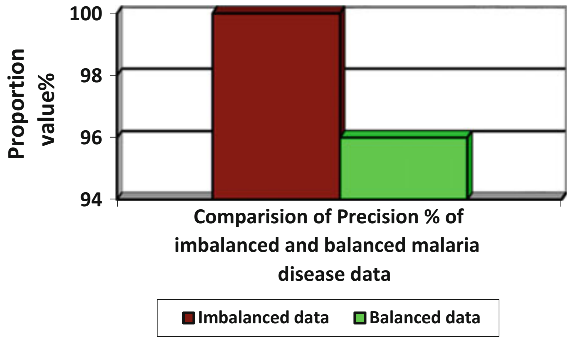


Fig. 5 Recall % of voting ensemble algorithm on both imbalanced and balanced malaria disease datasets

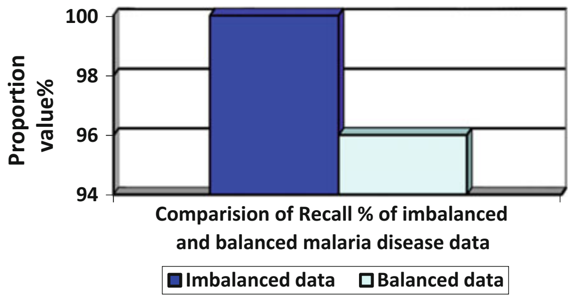
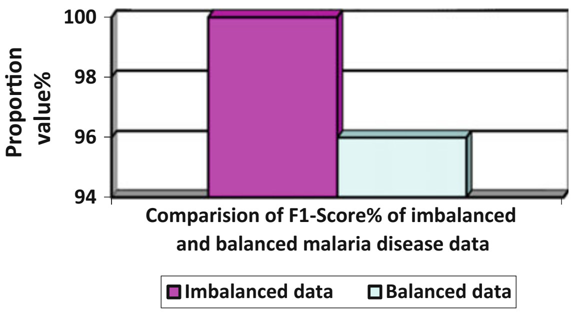


Fig. 6 F1-Score % of voting ensemble algorithm on both imbalanced and balanced malaria disease datasets



5 Conclusion

Vector borne diseases namely Malaria has high impact on world population today.

Even though vaccinations are providing still neonatals and children are effected by this disease. Hence proper diagnosing within stipulated time is very important other wise sometimes it causes to death also. So, accurate identification of patient and diagnosing within time plays very difficult task in medical domain. We are proposing an ensemble method called voting of C4.5, Naive Bayesian and KNN classifiers to classify the minority class samples especially for a better prediction system.

References

1. Bui TQ, Pham HM (2016) Web based GIS for spatial pattern detection: application to malaria incidence in Vietnam. Bui Pham Springer Plus 5(1014):1–14
2. MacLeod DA, Jones A, Di Giuseppe F, Caminade C, Morse AP (2015) Demonstration of successful malaria forecasts for Botswana using an operational seasonal climate model. Environ Res Lett 10:044005, 1–11 (IOP Publishing)
3. Rahman MZ, Roytman L, Kadik A, Rosy DA (2015) Environmental data analysis and remote sensing for early detection of dengue and malaria. In: Proceedings of SPIE, vol 9112, pp 1–9
4. WHO Malaria Report (2016) http://www.who.int/mediacentre/factsheets/fs387/en/
5. Pengfei J, Chunkai Z, Zhenyu H (2014) A new sampling approach for classification of imbalanced data sets with high density. In: IEEE—BigComp, pp 217–222
6. Ditzler G, Polikar R (2012) Incremental learning of concept drift from streaming imbalanced data. IEEE Trans Knowl Data Eng, pp 1–30
7. Nugroho HA, Akbar SA, Murhandarwati EEH (2015) Feature extraction and classification for detection malaria parasites in thin blood smear. In: IEEE 2nd international conference on information technology, computer and electrical engineering (ICITACEE), pp 197–201
8. Das DK, Maiti AK, Chakraborty C (2015) Automated system for characterization and classification of malaria-infected stages using light microscopic images of thin blood smears. J Microsc 257(3):238–252
9. Ruiz D, Brun C, Connor SJ, Omumbo JA, Lyon B, Thomson MC (2014) Testing a multi-malaria-model ensemble against 30 years of data in the Kenyan highlands. Malaria J 13:206, 1–14
10. Smith T, Ross A, Maire N, Chitnis N, Studer A, Hardy D, Brooks A, Penny M, Tanner M (2012) Ensemble modeling of the likely public health impact of pre-erythrocytic malaria vaccine. PLOS Med 9(1):1–20
11. Pandit P, Anand A (2016, August) Artificial neural networks for detection of malaria in RBCs. ArXiv: 1608.06627)
12. Bbosa F, Wesonga R, Jehopio P (2016) Clinical malaria diagnosis: rule based Classification statistical prototype. Springer Plus 5:939
13. Wu C, Wong PJY (2016) Multi-dimensional discrete Halanay inequalities and the global stability of the disease free equilibrium of a discrete delayed malaria model. Adv Differ Equ 2016:113
14. Tsai M-H, Tsai M-H, Yu S-S, Chan Y-K, Jen C-C (2015) Blood smear image based malaria parasite and infected-erythrocyte detection and segmentation. Transactional Processing Systems. J Med Syst 39:118. https://doi.org/10.1007/s10916-015-0280-9
15. Rahmanti FZ, Ningrum NK, Imania NK, Purnomo MH (2015, November) *Plasmodium vivax* classification from digitalization microscopic thick blood film using combination of second order statistical feature extraction and K-Nearest Neighbour (K-NN) classifier method. In: IEEE

4th international conference on instrumentation, communications, information technology, and biomedical engineering (ICICI-BME), Bandung, pp 2–3
16. Charpe KC, Bairagi V (2015) Automated malaria parasite and there stage detection in microscopic blood images. In: IEEE sponsored 9th international conference on intelligent systems and control (ISCO)
17. Somasekar J, Reddy BE (2015) Segmentation of erythrocytes infected with malaria parasites for the diagnosis using microscopy imaging. Comput Electr Eng, pp 336–351 (Elsevier)
18. Cameron E, Battle KE, Bhatt S, Weiss DJ, Bisanzio D, Mappin B, Dalrymple U, Hay SI, Smith DL, Griffin JT, Wenger EA, Eckhoff PA, Smith TA, Penny MA, Gething PW (2015) Defining the relationship between infection prevalence and clinical incidence of *Plasmodium falciparum* malaria. Nat Commun 6:8170, 1–10
19. Krawczyk B (2016) Learning from imbalanced data: open challenges and future directions. Prog Artif Intell, pp 1–12
20. Deng X, Zhong W, Ren J, Zeng D, Zhang H (2016) An imbalanced data classification method based on automatic clustering under-sampling. IEEE Trans, pp 1–8
21. Ali A, Shamsuddin SM, Ralescu AL (2013) Classification with class imbalance problem: a review. Int J Adv Soft Comput Appl 5(3):1–30
22. Poolsawad N, Kambhampati C, Cleland JGF (2014) Balancing class for performance of classification with a clinical dataset. In: Proceedings of the World Congress on engineering, vol 1, pp 1–6
23. Rahman MM, Davis DN (2013) Addressing the class imbalance problem in medical datasets. Int J Mach Learn Comput 3(2):224–228
24. Haixiang G, Yijing L, Shang J, Mingyun G, Yuanyue H, Bing G (2016) Learning from class-imbalanced data: review of methods and applications. Expert Syst Appl, pp 1–49
25. Jamal S, Periwal V, Scaria V (2013) Predictive modeling of anti-malarial molecules inhibiting apicoplast formation. BMC Bioinform 14:55, 1–8
26. Andrade BB, Reis-Filho A, Souza-Neto SM, Clarencio J, Carmargo LMA, Barral A, Barral-Netto M (2010) Severe *Plasmodium vivax* malaria exhibits marked inflammatory imbalance. Malaria J 9:13, 1–8
27. Dubey R, Zhou J, Wanga Y, Thompson PM, Ye J (2014) Analysis of sampling techniques for imbalanced data: An n = 648 ADNI study. Elsevier Neuro Image 87:220–241
28. Ng WWY, Hu J, Yeung DS, Yin S, Roli F (2015) Diversified sensitivity-based under sampling for imbalance classification problems. IEEE Trans Cybern, pp 1–11
29. Roumani YF, May JH, Strum DP, Vargas LG (2013) Classifying highly imbalanced ICU data. Health care Manag Sci 16:119–128
30. Pengfei J, Chunkai Z, Zhenyu H (2014) A new sampling approach for classification of imbalanced data sets with high density. In: IEEE transaction, pp 217–222
31. Garcia V, Sanchez JS, Mollineda RA (2012) On the effectiveness of preprocessing methods when dealing with different levels of class imbalance. Knowl Based Syst 25:13–21 (Elsevier)
32. Thongkam J, Xu G, Zhang Y, Huang F (2009) Toward breast cancer survivability prediction model through improving training space. Expert Syst Appl 36:12200–12209 (Elsevier)
33. Zhao X-M, Li X, Chen L, Aihara K (2007) Protein classification with imbalanced data. Wiley InterSci 70:125–1132
34. López V, Fernandez A, Garcia S, Palade V, Herrera F (2013) An insight into classification with imbalanced data: empirical results and current trends on using data intrinsic characteristics. Inf Sci 250:113–141 (Elsevier)
35. Ma L, Fan S (2017) CURE-SMOTE algorithm and hybrid algorithm for feature selection and parameter optimization based on random forests. BMC Bioinform 18:169

Automatic Detection of Malaria Parasites Using Unsupervised Techniques

Itishree Mohanty, P. A. Pattanaik and Tripti Swarnkar

Abstract The focus of this paper is towards comparing the computational paradigms of two unsupervised data reduction techniques, namely Auto encoder and Self-organizing Maps. The domain of inquiry in this paper is for automatic malaria identification from blood smear images, which has a great relevance in healthcare informatics and requires a good treatment for the patients. Extensive experiments are performed using the microscopically thick blood smear image datasets. Our results reveal that the deep-learning-based Auto encoder technique is better than the Self-organizing Maps technique in terms of accuracy of 87.5%. The Auto encoder technique is computationally efficient, which may further facilitate its malaria identification in the clinical routine.

1 Introduction

Malaria remains a life-threatening serious infectious disease and considered neglected, with roughly 200 million cases and 429,000 deaths per year according to the World Malaria Report 2016 [1]. The most severe and lethal form is caused by the genus Plasmodium that is transmitted and affects the red blood cells (RBCs) of the blood samples through the bites of infected female Anopheles mosquitoes. Various qualitative clinical assessments are being done, which involves manual counting of infected RBCs, microscopic diagnosis, and drug effectiveness in order to bring a control on the serious disease. The risk burden of being infected with malaria and developing the disease is still accounted due to lack of clinical and research exper-

I. Mohanty (✉) · P. A. Pattanaik · T. Swarnkar
Department of Computer Science & Engineering, S 'O' A (Deemed to be University), Bhubaneswar 751030, India
e-mail: itishreemohantym@gmail.com

P. A. Pattanaik
e-mail: priyadarshiniadyashapattanaik@soa.ac.in

T. Swarnkar
e-mail: triptiswarnakar@soa.ac.in

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_5

tizes, widespread of fake and substandard medicines, erroneous manual enumeration of malaria diagnosis, time-consuming visual diagnosis, handling of huge unlabelled data, lack of advanced machine learning tools and many more [2, 3].

Interestingly, many automated computer vision techniques have been developed for medical diagnosis that is useful for doctors to detect malaria parasite in microscopically thick blood smear images high magnification multi-views. These advanced techniques can help to improve diagnostic processing with less cost and reduce the clinical effort of pathologists. Considering, the above issues and approaches, we have proposed a quantitative comparative study of two unsupervised techniques for the diagnosis of malaria parasite cell samples using their microscopic blood smear high magnification images [4]. In this work, two types of the unsupervised Auto encoder (AE) and Self-organizing Maps (SOM) techniques are applied for segmentation of blood cells. The purpose of our paper is to handle the most challenging phase of malaria detection by segmenting the complex and varying shapes along with overlapping cells.

Figure 1 states the contribution of this paper where we present the latest systematic update SOM technique in image analysis and deep learning AE for malaria identification. The major contributions of our work are summarized as follows:

- We have developed a framework to compare between two unsupervised algorithms in terms of their computational performance, yielding to learn more robust and abstract semantic features for malaria parasite identification.
- Our experimental results show that our comparative learning framework can learn better deep representation features for diagnosis and improve the malaria prediction accuracy.
- The superior performance of deep-learning-based AE technique is one of the fast pre-processing algorithms, which has the ability to improve the image quality and helps in better feature representation using deterministic approach.

The remainder of this paper is summarized as follows: Sect. 2 describes the prior arts in the field of malaria parasite identification analysis; Sect. 3 describes the comparison study of two methodologies; Sect. 4 presents the details of the dataset used in the experimental setup and results in analysis, and Sect. 5 includes the conclusion of the comparative study of this work.

2 Prior Art

To our best knowledge, there are prior methods for diagnosing malaria from microscopic blood smear images [3], generally, follow processes including cell segmentation along with the extraction of useful features using computer vision classification techniques to classify infected and noninfected microscopic blood smear images. Das et al. [4] briefly describes the computational microscopic imaging methods on segmentation and feature extraction for malaria parasite identification. Jan et al. [5] reviews the complete overview of malaria diagnosis process of manual assessment

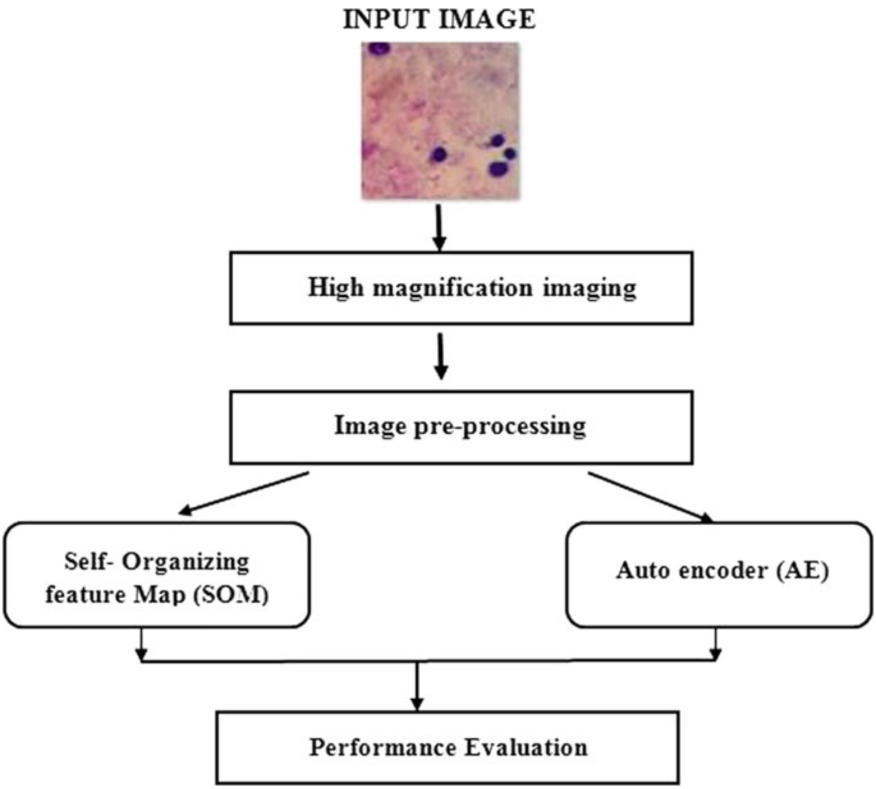


Fig. 1 Overview of our comparative paradigms of two unsupervised data reduction techniques, namely auto encoder and self-organizing maps to detect malaria parasite from microscopic blood smear images

and automated assessment in blood smear images. Devi et al. [6] states the different computational techniques used to detect the erythrocyte features for malaria detection in blood smear images. Poostchi et al. [7] describes the image analysis techniques used for imaging, image pre-processing, cell segmentation, and classification for detecting infected malaria slides for both thin and thick blood smear images. The paper discusses the latest developments in deep learning and smartphone technology towards malaria diagnosis. Shen et al. [8] proposes an image coding scheme by using the stacked autoencoder to achieve lossless compression for malaria infection diagnosis. The proposed deep learning autoencoder model provides a higher compression than other lossless methods like JPEG-LS, JPEG 2000 and CALIC. Bustamam et al. [9] state a clustering self-organizing map method (SOM) for papillomavirus detection which causes cervical cancer disease. Cervical cancer disease is one of the most dangerous diseases and SOM technique forms a clustering process for its identification with two of HPV in the first cluster and 16 others in the second cluster.

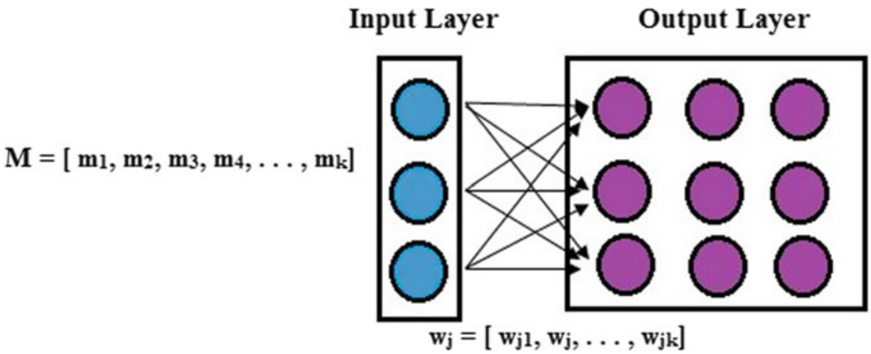


Fig. 2 Self-organizing feature map (SOM) structure

This paper describes the two unsupervised techniques been used in detail below.

3 Methods

The original microscopic blood smear malaria image dataset contains a significant amount of redundant information. However, to gain good computational quality of classification accuracy, image segmentation, and denoising have to be done in order to extract useful information of infected and noninfected blood components. Each color image is converted into a grayscale image through thresholding method and denoising is followed to improve the image quality. These are the data pre-processing steps to be carried followed by the two neural network techniques.

3.1 Self-organizing Map (SOM)

The Kohosen's self-organizing map is an unsupervised self-organizing artificial neural network composed of the input layer and output layers creating a static grid cell of fixed size. SOM technique forms a feedforward network structure with a computational layer in rows and columns from a higher dimensional input space to lower dimensional space [10]. The input layer carries input patterns to each individual nodes in the output layer through the weight matrix structure [11].

In Fig. 2, $M = [m_1, m_2, m_3, m_4, \ldots, m_k]$ is input vector and the weight vector that connect i-th node of the input layer and j-th node of output layer w_{ij} ($j = 1, 2, 3, \ldots, S$) is followed by

$$w_j = \left[w_{j1}, w_j, \ldots, w_{jk}\right] \tag{1}$$

The weights on the input neurons are randomly initialized and the weights are updated by two basic methods. The first method (Best matching unit takes all) includes the neuron whose weights are close to the input vector components gets modified in such a way that the weights get updated as possible to the input vector. The second method (Best matching unit takes most) includes the neuron whose weight is same as the input value gets the most priority. This way its weights and neighboring neurons weight are updated for a given number of epochs or satisfy the given stop criteria.

The update methods of the weight vector $w_q(t)$ at time t is,

$$w_q(t+1) = w_q(t) + Þ(t)\,(M - w_q(t)) \tag{2}$$

where $Þ(t)$ learning coefficient value at time t.

3.2 Autoencoder

Autoencoder is a typical fully connected encoder-decoder architecture [8, 12] where it reconstructs its input vector 'M' at the output 'S' using deterministic approach. Let us consider a set of data samples $M = [m_1, m_2, m_3, m_4, \ldots, m_k]$ in the input vector where $m_i \in R^d$, the training objective of an autoencoder is to minimize the reconstruction error.

$$1 = \sum_k ||mk - mk^{\because}||^2 \tag{3}$$

where $\mathrm{m_k}$ and $m_k^{\because}$ are the input vectors and the hidden layer includes both the encoding and decoding processes of

$$\begin{cases} h_k = f(W_{mk} + b) \\ m_k^{\because} = f\left(W' h_k + b'\right) \end{cases}, \tag{4}$$

where $h_k \in R^n$ as the compact representation and W, W', b, b' as the weights of encoding layer, decoding layer, and the bias terms respectively. The sigmoidal function in this paper can be represented using f (.) activation function,

$$f(o) = \frac{1}{1 + \exp(-o)} \tag{5}$$

4 Experimental Setup

To demonstrate the effectiveness and importance of the comparative model, we conducted experiments using the publicly available 1182 thick blood smear malaria images dataset consisting of 750×750 pixels obtained from Android smartphone to a Brunel SP150 microscope from Makerere University. The images were randomly splitted into training and testing images collected from different patients and the full dataset is provided in this link: http://air.ug/index.html.

4.1 Evaluation Criteria

To evaluate the comparative techniques, quantitative evaluation criteria includes calculation of accuracy, sensitivity, and specificity using various proportions of true results. Accuracy is the degree of veracity proposition test of true outcomes from either true positive (TP) or true negative (TN), in the samples. The performance measure calculation is illustrated

$$\text{Accuracy} = (\text{TP} + \text{TN})/(\text{TP} + \text{TN} + \text{FP} + \text{FN}) \quad (6)$$

$$\text{Sensitivity} = (\text{TP})/(\text{TP} + \text{FN}) \quad (7)$$

$$\text{Specificity} = (\text{TN})/(\text{TN} + \text{FP}) \quad (8)$$

where TP = True Positive, TN = True Negative, FP = False Positive, FN = False Negative.

4.2 Results and Discussion

In this comparative paradigm, we presented the experiments with a patient level by randomly dividing the whole dataset into 70% training and rest 30% in testing. As per Fig. 1, this approach of using blood smear image cytological criteria for diagnosis involves (i) SOM technique used to segment the cell components present in blood smear images. (ii) Quantification and deep extraction of cell features are done using deep learning technique named Autoencoder. Various morphometric features from each of the segmented blood smear image includes color histogram, color autocorrelogram, area granulometry, gradient descent, relative shape measurements, hu moments, scale invariance, number of colors of diagnosis the malaria parasite infection in erythrocyte, nuclear density, nucleocytoplasmic, Euler number are, fractal dimension and color channel histogram, phase of image, skewness, kurtosis, standard deviation, Sobel histogram, flat texture, co-occurrence matrix and run-length matrix.

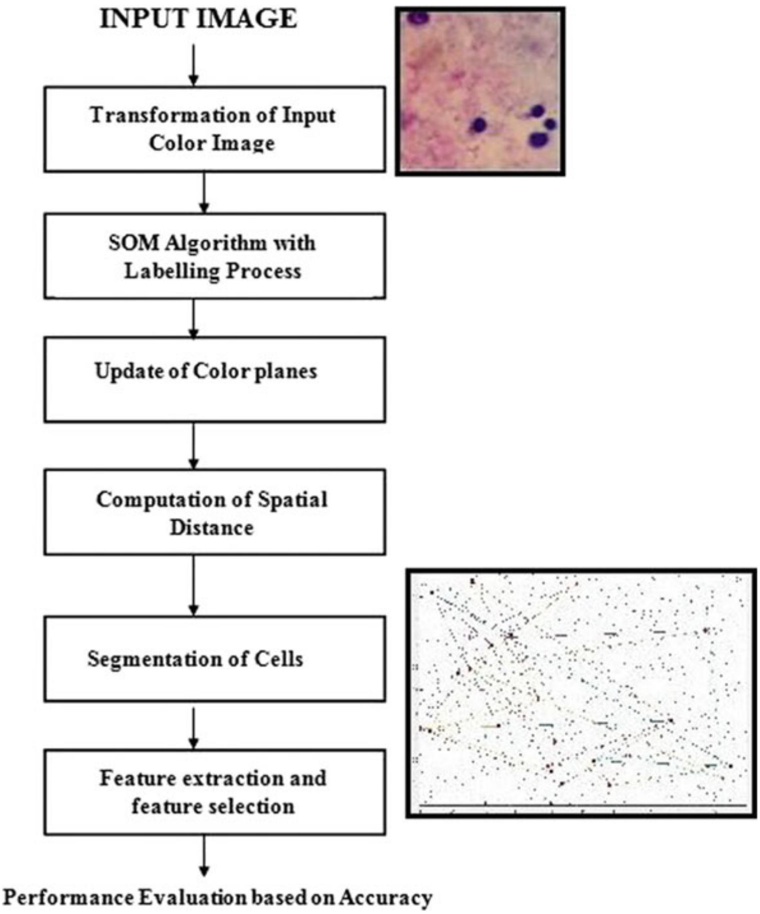


Fig. 3 Conceptual illustration for identifying malaria in the blood smear image using self-organizing feature map (SOM) technique

In SOM technique, each identified blood smear blood cell is segregated from other neighboring blood cells based on minimum area. To optimize the technique performance, we have extracted the malaria features from already computed SOM technique. For blood component classification, we have used texture, color, and shape as the important features for differentiating infected and noninfected blood smear images. The healthy blood smear images contain no nucleus in RBCs whereas unhealthy blood smear images contain the central pale area with some abnormalities in cells as shown in Fig. 3. Using the SOM technique, the spatial distance between blood components is calculated and performance based on accuracy is measured. This technique delivers an accuracy of 0.79 with a sensitivity of 0.80 and specificity of 0.78 in detecting malaria-infected blood smear images (Fig. 4).

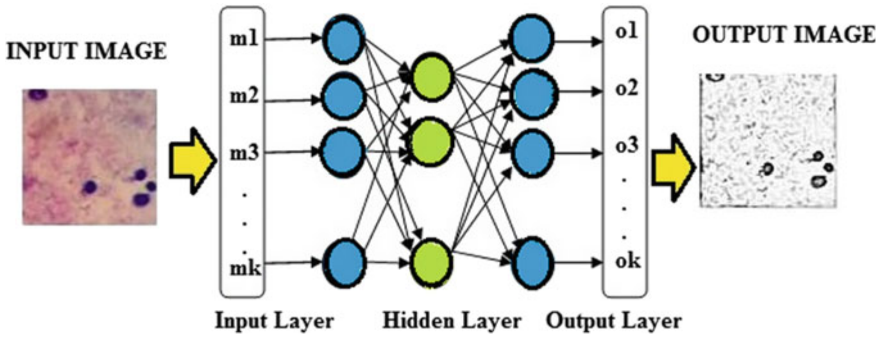


Fig. 4 Autoencoder (AE) structure

For simulation purpose, we have used MATLAB 2016B: an open source library. In this section, we have performed on a workstation with Intel Core i5 CPU and 16 GB of memory. The input to the Auto encoder technique is a mat file of 1182 images of pixel intensities, which further gets converted to vector input layers after layers. The autoencoder model is an unsupervised pre-trained model with a learning rate of 0.05 and the number of epochs set to be 1000 [8]. In the first layer or input layer, AE model generates 12,500 input nodes and in the output section of the decoder layer, 30 useful features were noted. The models are all trained from scratch and are running with an epoch of 1000. Through 1000 epochs, 30 useful features were considered for representing infected and noninfected blood smear images. Considering the input nodes and output nodes, performance assessment based on accuracy is done and explained further. This AE technique delivers an accuracy of 0.875 with a sensitivity of 0.84 and specificity of 0.80 in detecting malaria-infected blood smear images.

5 Conclusion

In this work, we have presented an efficient comparative approach based on malaria identification that focuses on the benefits of using two unsupervised neural network based models. The comparative models learn robust and hierarchical representations from unlabelled data and their experimental results show the effectiveness in identifying malaria parasite in blood smear images. Table 1 presents classification accuracy performance of both the techniques along with the details of training and test data in each experiment.

The blood smear based cell sample classification achieves the accuracy of 79% using SOM technique which is less than the classification accuracy of 87.5% using AE technique. This shows the comparative framework ability to identify the infected cells as well as noninfected cells in blood smear images.

Table 1 Accurancies of the two different neural network techniques to identify malaria in blood smear images

Techniques	Accuracy	Sensitivity	Specificity
SOM technique	0.79	0.80	0.78
Auto encoder	0.875	0.84	0.80

References

1. WHO (2016) Malaria microscopy quality assurance manual-version 2. World Health Organization
2. WHO (2016) World malaria report 2016. World Health Organization
3. Tek FB, Dempster AG, Kale I (2009) Computer vision for microscopy diagnosis of malaria. Malaria J 8(1):153
4. Das D, Mukherjee R, Chakraborty C (2015) Computational microscopic imaging for malaria parasite detection: a systematic review. J Microsc 260(1):1–19
5. Jan Z, Khan A, Sajjad M, Muhammad K, Rho S, Mehmood I (2017) A review on automated diagnosis of malaria parasite in microscopic blood smears images. Multimedia Tools Appl 77:1–26
6. Devi SS, Sheikh SA, Laskar RH (2016) Erythrocyte features for malaria parasite detection in microscopic images of thin blood smear: a review. Int J Interact Multimed Artif Intell 4(2):34–39
7. Poostchi M, Silamut K, Maude R, Jaeger S, Thoma G (2018) Image analysis and machine learning for detecting malaria. Transl Res 194
8. Shen H, Pan WD, Dong Y, Alim M (2016) Lossless compression of curated erythrocyte images using deep autoencoders for malaria infection diagnosis. In: Picture Coding Symposium (PCS), pp 1–5
9. Bustamam A, Aldila D, Fatimah, Arimbi MD (2017) Clustering self-organizing maps (SOM) method for human papillomavirus (HPV) DNA as the main cause of cervical cancer disease. In: AIP conference proceedings, vol 1862, no 1, pp 30–155
10. Corral JA, Guerrero M, Zufiria PJ (1994) Image compression via optimal vector quantization: a comparison between SOM, LBG and *k*-means algorithms. In: 1994 IEEE international conference on neural network. IEEE World Congress on computational intelligence, vol 6, pp 4113–4118
11. Marghescu D, Rajanen MJ (2005) Assessing the USE of the SOM technique in data mining. In: Databases and applications, pp 181–186
12. Razzak MI, Naz S, Zaib A (2018) Deep learning for medical image processing: overview, challenges and the future. In: Classification in BioApps, pp 323–350

Study on Different Region-Based Object Detection Models Applied to Live Video Stream and Images Using Deep Learning

Jyothi Shetty and Pawan S. Jogi

Abstract There is a plenty of very interesting problems in the field of computer vision, from the very basic image classification problem to 3d pose estimation problem. One among the many interesting problems is object detection, which is the computer capability to accurately identify the multiple objects present in the scene (image or video) with the bounding boxes around them and the appropriate labels indicating their class along with the confidence score indicating the degree of closeness with the class. In this work, we have discussed in detail different types of region-based object detection models applied on both live video stream and images.

1 Introduction

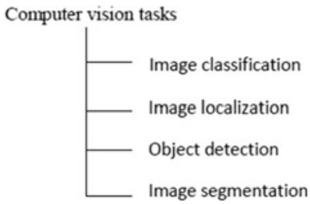
Object detection has been one of the hot topics in the area of computer vision. Computer vision aims at enabling vision (ability to see) capability to the machines. Different tasks under computer vision includes (1) Image classification which basically involves classifying the given input image accurately to one of the many classes with the confidence score indicating the degree of closeness with the class. (2) Localization is similar to classification but in addition, it involves identifying the exact location of the objects present in the image with the bounding box. (3) Object detection is one of the major tasks under computer vision, which involves localizing distinct objects present in the image. (4) Image segmentation is the next stage of object detection, which performs a pixel-by-pixel segmentation of multiple objects present in the query image. Figure 1 shows different tasks under computer vision. The emergence of deep learning has completely changed the traditional methods of performing computer vision tasks. In 2012 Alex Krizhevsky, Geoffrey Hinton et al. presented a paper "ImageNet classification with deep convolutional neural network", a convolutional

J. Shetty · P. S. Jogi (✉)
NMAMIT, Nitte, Karkala, India
e-mail: pawansnj03@gmail.com

J. Shetty
e-mail: jyothi_shetty@nitte.edu.in

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_6

Fig. 1 Different computer vision tasks



neural network (CNN or ConvNet) based deep learning model which reached the state of the art in the task of performing object detection and won the annual ImageNet challenge (International event on performing computer vision tasks), dropping the error rate from 26 to 15%. This was the first ever successful deep learning model.

Neural networks played a remarkable role in the area of computer vision. A model named ConvNet is performing extremely well in the field of computer vision. ConvNets are the biologically inspired models based on the theory proposed by the researchers D. H. Hubel and T. N. Wissel, motivated by the detailed study of the working of human brain cells called neurons (the exact working of a human brain is still an unsolved problem). They demonstrated how the human brain would interpret images or videos by the simulation of multiple layers of billions of clusters of interconnected neurons, which are responsible for getting a more abstract view of the objects, this turned the researchers to develop the similar pattern recognition models for computer vision tasks. Object detection finds its application in ADAS (Advanced Driver Assistance System), advanced robotics (Humanoids), helping visually challenged people, surveillance system, pedestrian detection, face recognition, military system, healthcare, security system, visual search, pose recognition, gaming, space research, and many more.

The paper is structured as follows. In Sect. 2, we have discussed the evolution of region-based object detection techniques, Sect. 3 is about a detailed study on different object detection models applied on live videos and images, Sect. 4 draws the conclusion.

2 Region-Based Approaches for Object Detection Models

The evolution of region-based approaches changed the direction of performing computer vision tasks. Ross Girshreik (researcher at Facebook AI) and his team at UC Berkeley invented one of the most effective models for performing computer vision tasks. The objective is very simple, for an image the model must be able to identify all the objects present in it by putting the bounding boxes around them along with the appropriate labels. There are two types [1] (1) Single-stage object detection model and (2) Two-stage object detection model. Figure 2 shows different region based convolutional neural network models.

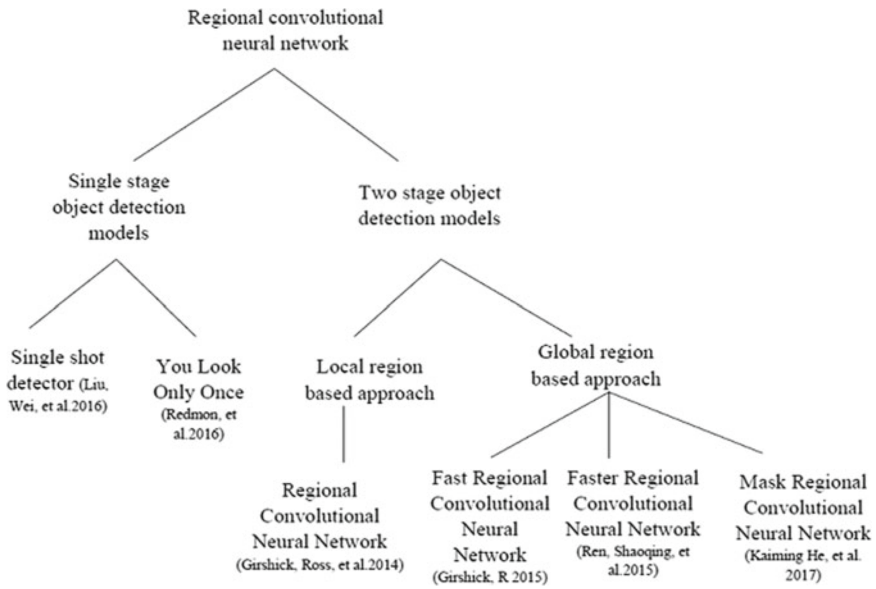


Fig. 2 Different types of region-based convolutional neural network

2.1 Single-Stage Object Detection Models

Single-stage object detection models are the feed-forward neural network, which can perform object detection. These kind of models are not using any kind of object proposal generating methods (techniques for identifying the different parts of the image which probably contain the objects) and resampling stages, instead, all these stages are combined in a single stage so that the computation cost can be reduced which will increase the performance. The whole purpose of single-stage object detection model is speed, however, they lack accuracy. Some of the examples of single-stage object detection models are Single-Shot Detector (SSD) and You Look Only Once (YOLO).

Single-Shot Detector Model (SSD) Single-Shot Detector (SSD) [2] is a single pipelined deep neural network model for performing object detection. This model is having the capacity to combine the predictions generated from the different feature maps with a variable resolution to handle objects of distinct size. SSD gets rid of separate steps for performing object proposal generation and resampling, instead, it binds all these in a single framework. Elimination of the object proposal generation resulted in the decrease in computation cost and increase in performance to the greater extent. This model can be directly integrated with other systems that require object detection module. It is faster but less accurate when compared with the two-stage object detection models. SSD performed well on datasets like MSCOCO, PASCAL VOC.

You Look Only Once Model (YOLO) You Look Only Once (YOLO) [3] is a single deep neural network, which is capable of predicting the bounding boxes for the objects and finding out the probability of the class to which it belongs from the whole image, since the entire process involves only one network, end-to-end optimization is possible. YOLO is very fast when compared with any other object detection models, the base YOLO model is capable of processing 45 frames per second, whereas Fast YOLO processes 150 frames per second. YOLO is capable of processing real-time streaming videos with less than 25 ms of latency. This model produced competitive accuracy as two-stage object detection models. YOLO detection mechanism involves three stages, (1) Resizing the query image to 448 $\times$ 448 pixels. (2) Running single convolution neural network. (3) Classifying the objects by using model's confidence. YOLO performed well on VOC dataset. However, YOLO failed to reach the accuracy level of two-stage object detection models.

2.2 Two-Stage Object Detection Models

The first stage of the two-stage detection model is using object proposal generation methods (techniques for identifying different regions of the image which probably contain the objects) followed by decision refinement. Some well-known methods like regional convolutional neural network used selective search to generate object proposals, and ConvNets to extract and feed features to the classifier. The second phase involves refining the boundary boxes, identifying the fake detections and rescoring the boxes based on the other objects detected on the image. This is a very complex process and is very difficult to optimize since each of the individual components must be trained independently. Two-stage object detection models are having very high accuracy but very difficult and slow to train when compared with the single-stage object detection models.

Regional Convolutional Neural Network Model (R-CNN) Regional convolution neural network (R-CNN) [4] is a local region-based convolutional neural network. This model is based on identifying the different regions that could possibly contain object followed by the running classifier in order to identify the different objects. This method combines two key features (1) Using high-capacity ConvNets to identify the objects and (2) Supervised learning for the objects in a case when there is a shortage of the labeled training data to get efficient results. This system (1) Takes the query image of size 227 $\times$ 227 pixels (2) extracting approximately 2000 bottom-up region proposals by using selective search (3) finding the features for each of the proposal (bypassing the local regions of the image) using large CNN's, and then (4) classification is performed by using linear support vector machine (LSVM). This was a very complex task and the performance of the R-CNN was very poor because it should perform convolution for every region. R-CNN produced good results on ILSVRC 2013 dataset. The main drawback of R-CNN was its low speed. Some of the disadvantages of R-CNN model are listed below. (1) Training involves multiple stages (2) Training is not cost efficient. (3) Object detection is slow.

Fast Regional Convolutional Neural Network Model (Fast R-CNN) Fast regional convolution neural network (Fast R-CNN) [5] is the next version of the original R-CNN. Fast R-CNN consumes less computation time and improves the detection accuracy, due to the following reasons (1) Feature extraction is performed on the image before generating the object proposals thus it will be enough running one CNN over the entire image (passing the global or full image for feature extraction) instead of running 2000 ConvNets over 2000 proposed regions. (2) Using softmax instead of SVM for classification. Then for every object, a pooling layer called Region of Interest (RoI) will extract a fixed-length feature vector from the feature map. Thus, obtained feature vectors are fed onto a series of fully connected layers. This made the model not only fast, but also end-to-end trainable. In Fast R-CNN the object proposal generation algorithms like selective search (SS) played a very crucial role (even though it was less efficient). Advantages of the proposed model are (1) Higher detection accuracy. (2) Training is a single stage. Fast R-CNN reached the state of the art on VOC07, 2010, and 2012.

Faster Regional Convolutional Neural Network Model (Faster R-CNN) Faster regional convolution neural network (Faster R-CNN) [6] is the part 3 of R-CNN series. The whole idea of this model is to replace the selective search algorithm (one of the major drawbacks of Fast R-CNN is its low speed) and to make the model end-to-end trainable. The core idea behind Faster R-CNN model was that region proposals depended on the features of the image that were already calculated with the forward pass of the ConvNets. They have used the same ConvNets results for region proposal generation instead of running selective search algorithm. Region proposal network (RPNs) were introduced. RPN is a fully convolution neural network on top of the features of the ConvNets, which is capable of generating the bounding boxes for the object along with the scores indicating the degree of closeness with the class to which it belongs. The proposed system was cost efficient and produced very high accuracy on PASCAL VOC dataset.

Mask Regional Convolutional Regional Neural Network (Mask R-CNN) Mask regional convolutional neural network (Mask R-CNN) [7] aimed at performing pixel level segmentation. Mask R-CNN is built on top of the Faster R-CNN with the additional step called mask. It takes a pixel and predicts whether it s a part of the object or not. Mask is a fully convolutional neural network built on top of the ConvNet based feature map. Workflow of mask R-CNN is very simple it will take CNN feature map as the input and outputs the matrix with 1's in all location in case the pixel belongs to the same object or 0 otherwise. Mask R-CNN performed well on MSCOCO dataset.

3 Experiments and Results

Creating the deep learning models with the ability to localize and identify the multiple objects present in the image with the highest accuracy is considered as one of the challenging problems to succeed. Google released object detection framework which is built on top of tensorflow (deep learning library released by Google Brain Team). This helped researchers and developers to build their own custom object detection models. It allows us to use pretrained models for performing computer vision tasks. These models are having mAPs (mean average precision), which tell us how sensitive the model is for objects of interest and how well the model handles the fake detections. As the mAP value increases the models will tend to have more accuracy, higher the accuracy the cost of execution increases. These models are trained on MSCOCO dataset. MSCOCO (Microsoft Common Object in Context) is one of the most widely used datasets for performing computer vision tasks. This dataset contains approximately 330,000 images along with associated labels, which belong to 90 of the most commonly found objects.

3.1 Object

In this experiment, we have detection on images performed the object detection using SSD Mobilenet V1 COCO (single-stage object detection model) and Faster R-CNN Resnet101 COCO (two-stage object detection model) and applied it on both images and live video stream. We found that SSD was able to output the result in less time with low accuracy having only one bounding box around the objects, whereas Faster R-CNN model generated results with more accuracy by putting multiple bounding boxes on objects but it took more time than SSD. We have shown results in Fig. 3.

3.2 Object Detection on Live Video Stream

We have performed object detection on a live video stream by using TensorFlow and OpenCV. OpenCV is a library for processing the images as well as videos. It supports C++, C, Python, and Java. OpenCV is used for performing all kind of video and image analysis like face recognition and detection, license plate reading, advanced robot-vision, and many more. We have performed this experiment by using SSD and Faster R-CNN models results are shown in Fig. 4, SSD model was unable to identify all the objects present in the image, whereas Faster R-CNN model accurately identifies all the objects present in the image.

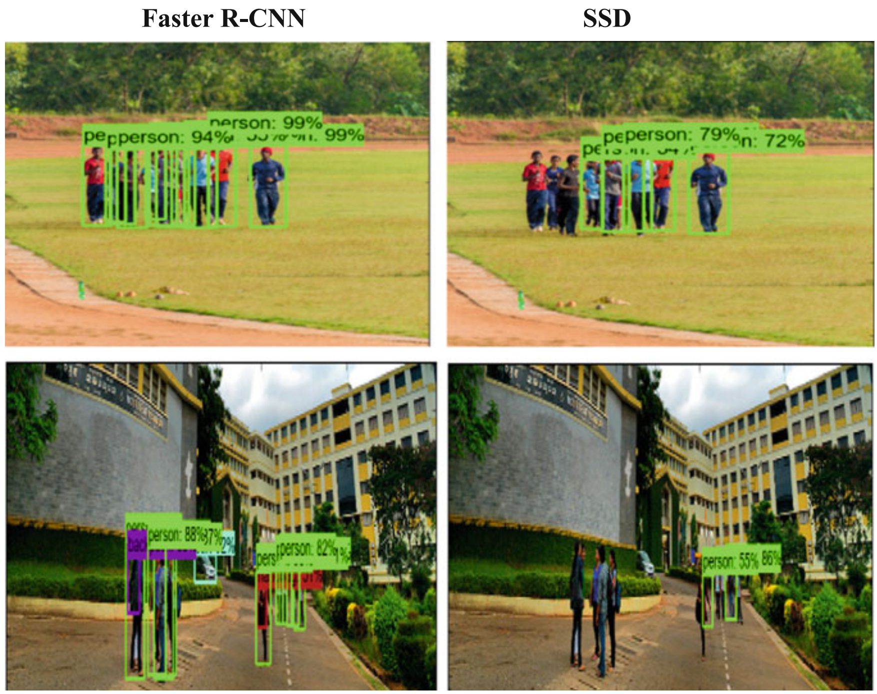


Fig. 3 Comparison of SSD and faster R-CNN models

Fig. 4 Comparison of SSD and faster R-CNN model in the live video

3.3 Images Versus Time Comparison

In this experiment, we have taken a total of 500 images (in 100 intervals), all images are <1 mb and performed object detection using SSD and Faster R-CNN models. As expected SSD was able to produce the results quickly with less accuracy whereas,

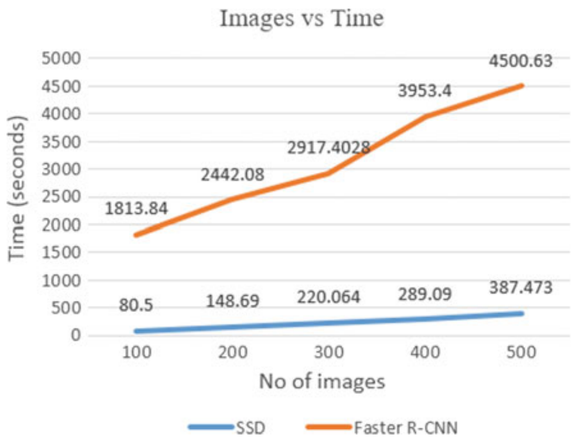


Fig. 5 Number of images versus time graph

Faster R-CNN was able to produce results which are highly accurate but it took very long time when compared with SSD. We plotted images versus time graph. Results are shown in Fig. 5. We also found that as the image size increases the models took more time to output the result.

3.4 Accuracy Comparison Between SSD and Faster R-CNN

Accuracy is a very important factor for any machine learning models. Here we have investigated the accuracy level of different convolutional neural network based object detection models. We have two types of object detection models one focusing on accuracy (Two-stage object detection models. Ex: Faster R-CNN) and another one on speed (Single-stage object detection models. Ex: SSD). So, it is very important to choose the model which best suits our application. We have evaluated the accuracy based on the number of accurately identified objects among the total number of objects present in the image. We have examined that, as the accuracy level of the model increases the speed will decrease drastically [8]. So it is very clear that, designing a model which is good at both accuracy and speed is very hard. We have shown our results on live videos in Fig. 6, followed by on images in Fig. 7. We have performed all the experiments on Ubuntu 16.04 LTS core i3 processor with 8 GB RAM and 1 TB hard disk having a processor speed of 2.00 GHz and tensorflow 1.4.

4 Conclusion

In this work, we have discussed in detail about the different types of region-based object detection models. There exists a trade-off between speed and accuracy among

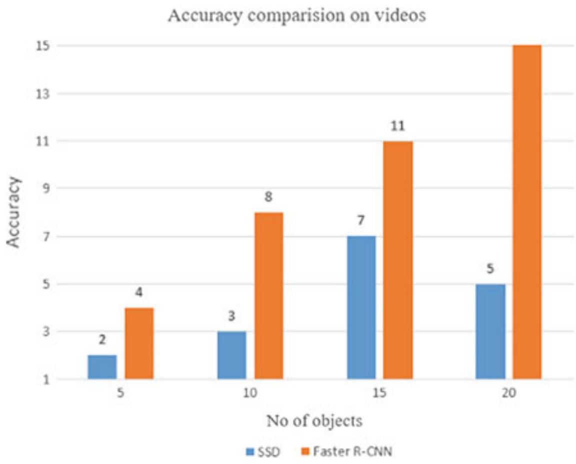


Fig. 6 Accuracy comparison on videos

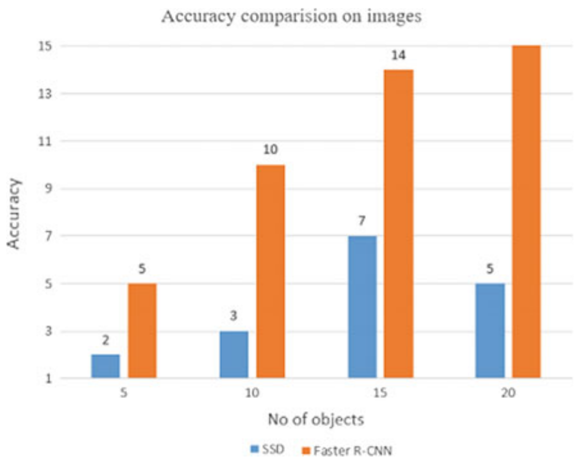


Fig. 7 Accuracy comparison on images

different types of deep learning based object detection models. This made us to investigate about the speed and accuracy of various object detection algorithms on both images as well as on live videos which will help researchers and developers to choose the detection architecture depending upon their requirement.

Acknowledgements We are thankful to NMAM Institute of Technology and Engineering Nitte, for allowing us to take the images required for performing the experiment.

References

1. Huang J, Wang D, Wang X (2017) Novel single stage detectors for object detection
2. Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu CY, Berg AC (2016) Ssd: single shot multibox detector. In: European conference on computer vision. Springer, Cham, pp 21–37
3. Redmon J, Divvala S, Girshick R, Farhadi A (2016) You only look once: unified, real-time object detection. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 779–788
4. Girshick R, Donahue J, Darrell T, Malik J (2014) Rich feature hierarchies for accurate object detection and semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 580–587
5. Girshick R (2015) Fast r-cnn. In: Proceedings of the IEEE international conference on computer vision, pp 1440–1448
6. Ren S, He K, Girshick R, Sun J. (2015) Faster R-CNN: towards real-time object detection with region proposal networks. In: Advances in neural information processing systems, pp 91–99
7. He K, Gkioxari G, Dollár P, Girshick R (2017) Mask r-cnn. arXiv preprint arXiv:1703.06870
8. Huang J et al (2017) Speed/accuracy trade-offs for modern convolutional object detectors. IEEE CVPR

Digital Video Copy Detection Using Steganography Frame Based Fusion Techniques

P. Karthika and P. Vidhyasaraswathi

Abstract An effective and exact technique for copying video location in a huge dataset is the utilization of video pictures. We have exactly picked the shading format description, a minimal and powerful casing-based description to make pictures which are additionally determined by vector quantization (VQ). We recommend a new nonmetric length measure to discover the likeness among the question and a dataset video picture and tentatively demonstrate its better execution over other length measures for exact copy identification. The effective look cannot be executed for high-dimensional information utilizing a nonmetric distance measure with accessible ordering systems. Consequently, we create a novel search algorithm in the view of precompiled distances and new dataset reduce systems yielding reduced recovery times. We perform different things with colossal dataset recordings. For singular questions with a normal span of 60 s (around half of the normal dataset video duration), the copy videos are recovered in 0.032 s, on Intel Xeon with CPU 2.33 GHz, with a high exactness of 98.5%.

1 Introduction

The productions of the digital image are an essential and test errand in shared correspondence and different strategies are utilized to produce advanced pictures. There are three objectives of system or data security, for example, secrecy, uprightness, and accessibility (CIA). Privacy implies that data is secure and not accessible to the unapproved individual. Honesty alludes to the precision of data and accessibility implies that data is in accessing time to the approved individual. Picture protection is not adequate for dependable data like content, sound, video, and computerized pictures.

P. Karthika (✉) · P. Vidhyasaraswathi
Department of Computer Applications, Kalasalingam Academy of Research and Education, Krishnan Koil, Tamil Nadu, India
e-mail: karthikasivamr@gmail.com

P. Vidhyasaraswathi
e-mail: vidhyasaraswathi.p@gmail.com

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_7

There are numerous methods to secure pictures including watermarking, advanced watermarking, reversible watermarking, cryptography, steganography, and so on. In this paper, video duplicate identification utilizing steganography is described. We proposed a half and half production of data to approach that is a combination of steganography and frame-based fusion techniques. A short presentation of every procedure has been discussed in the following sections.

A. **Steganography**

Invisible communication has been possible through the steganography. In steganography, the first image is hidden within the cowl image to masquerade the trespasser or hacker and therefore the resulted image is termed stego image. To discover the hidden pictures, a corporation should actively monitor network traffic that is time and processor intensive. However, those that are at home with the network's traditional traffic patterns will merely rummage around for changes like the accumulated movement of enormous pictures across the network, which can warrant additional, careful investigation. It is also informed to have and actively enforce a security policy that clearly outlines acceptable usage, what information will and cannot be sent across the network and the way it should be protected. Also, limit unauthorized programs, the utilization of unauthorized encoding and steganography within the geographical point and think about limiting the scale of mailboxes.

B. **Video copy detection**

It is the method of work lawlessly traced videos by analyzing and examination them to original content today the unlawful issues making on audio, video and text into numerous forms through the net. The electronic communication net is everything is visible and accessible to each user. Therefore, image security of knowledge could be a necessary and vital task varied techniques victimization the copy video detection for steganography is that the art of concealment the very fact communication is going down, by concealment data in different data many various carrier file formats may be used, however digital pictures area unit the foremost in style thanks to their frequency on the net. This paper introduces two new ways, wherever one is fusion and another one is steganography area unit combined to write in code the info still on hide the info in another medium through image process.

C. **Fusion Techniques**

The image combinations of the extraordinary information from every particular picture are melded structure. A consequential excellence image is healthier than any knowledge of the pictures. Image fusion technique will be roughly categorized into two.

1. Abstraction area Fusion Technique
2. Spatial and Temporal Domain Fusion Technique

Abstraction area Fusion Technique: A pair of remodel domain fusion and abstraction domain techniques, we have a tendency to specifically manage the image pixels values square measure controlled to accomplish made-up outcomes. In return house techniques, the image is first moved into the return space.

Spatial and Temporal Domain Fusion Technique: We have a tendency to reduce image pixels and manage them to get the specified yield. In remodel Domain technique and Fourier Transformation, square measure is applied to the image to urge the specified yield. A number of the primitive ways in image fusion square measure averaging technique is to pick the most and minimum. It is a derived composition generated from a sequence of pictures and may be obtained by understanding geometric relations between the photographs.

D. **Context Analysis for Retrieval**

Relevant data has been effectively talked about from various perspectives, going from the spatial, transient, shape setting, design, and relevant setting. Labels and areas are two normally utilized setting data for picture recovery. Social connections have pulled in the considerations which are utilized to contemplate the client-to-client, and client-to-photograph relations. The client setting and interpersonal organization setting were additionally used to comment on pictures. In some current investigations, the combination of substance and setting is a typical method to enhance the execution. The semantic setting prompted from the discourse transcript encompassing a key casing was joined with the visual catchphrases to enhance the execution of close copy picture recovery. A setting chart was developed at the record point for video seek reranking, in which setting alludes to the properties depicting who, where, when, what, and so forth, of web archives. Labels, clarification, migrations, and visual information have been used to enhance the recovery data for Flickr. In labels, area and substance investigation (shading, surface, and neighborhood focus) were utilized to recover the pictures of topographical-related milestones. The vast majority of the specified works are basically in light of the picture sharing site Flickr. Be that as it may, there has been small research investigating the setting data for video sharing sites, for example, YouTube. It stays vague whether relevant assets are likewise successful for web recordings. Specifically, the combination of substance and setting data for close copy web video end has not been genuinely tended to.

2 Literature Survey

Video duplicate and comparability discovery have been effectively considered for its potential to seeking [1], subject following [2], and copyright security [1]. Different methodologies utilizing distinctive highlights and coordinating calculations have been proposed. Among existing methodologies, numerous accentuate the quick distinguishing proof of copy recordings with smaller and solid worldwide highlights. These highlights are by and large alluded to as marks or fingerprints which outline the worldwide measurement of small level highlights. Run of the mill highlights incorporate shading, movement and ordinal mark [3, 4] and model-based mark [1, 5]. These worldwide component-based methodologies are appropriate for recognizing relatively indistinguishable recordings, and can identify minor altering in the spatial and fleeting space [1, 3–5, 9]. Our examination of an assorted arrangement of well-known web recordings demonstrates that there are around 20% correct copy recordings among all close copy web recordings [6]. It is normal for web clients to transfer correct copy recordings with insignificant change. This features the need of an approach for quick location of copy recordings. Be that as it may, world-wide highlights end up plainly inadequate when managing video duplicates covered in layers of altering beauty care products. For these more troublesome gatherings, low-level highlights at the fragment or attempt level are useful to encourage nearby coordinating [4, 6–8, 9]. Commonly, the granularity of the fragment level coordinating, the adjustments in the worldly request, and the addition/erasure of edges all add to the similitude score of recordings. Contrasted with signature-based techniques, portion-level methodologies are slower, however, fit for recovering estimated duplicates that have experienced a generous level of altering. At a larger amount of multifaceted nature, more troublesome copies with modification in foundation, shading, and lighting, need considerably more perplexing and solid highlights at district stage. Highlights, for example, shading, surface, and shape can be separated at the key casing level, which thus could be additionally portioned into various area units. Notwithstanding, the issue of division unwavering quality and the granularity choice brings into question the adequacy of these methodologies. As of late, nearby intrigue focuses (key focuses) are appeared to be helpful for close copy and duplicate recognition [1, 3–6, 41]. Nearby focuses are striking neighborhood areas recognized over picture scales, which find neighborhood locales that are open-minded to geometric and photometric varieties [7]. Striking areas in each key edge can be separated with neighborhood point indicators and their descriptors are for the most part invariant to nearby changes. Keypoint based nearby element identification approach stays away from the deficiency of worldwide and portion level highlights, and gives a precise estimation notwithstanding for pictures experienced extreme changes. Albeit neighborhood focuses have generally been recognized as solid and vigorous highlights, productively coordinating extensive measure of nearby focuses remains a troublesome issue. Late arrangements incorporate utilizing ordering structure and quick close copy following with heuristics [1]. On a very basic level, the errand of close copy discovery includes the estimation of repetition and curiosity, which has been investigated in content data recovery [2, 7]. The curiosity recognition approaches for records and sentences chiefly center around vector space replica and factual dialect models to quantify the level of oddity communicated in words. Question pertinence and data oddity have been consolidated to re-rank the reports/pages by utilizing graph [2] and dialect models [3]. To the most excellent of our insight, there is little research on close copy video recognition and re-positioning for substantial scale web video investigates [1].

3 Problem Analysis

1. Local element based CBCD calculations have demonstrated better recognition rate, and however, extraction of nearby highlights alongside their coordinating procedure have noteworthy time prerequisites.
2. Video preprocessing done by CBCD frameworks incorporates expulsion of dark outskirt, picture-in-picture, and camcording impacts. Because of such preprocessing, the worldwide highlights can adequately manage intense changes.
3. Applied idea of sack with DCT-sign based component, which is typically utilized with nearby highlights. Accordingly to applying the ideas of nearby highlights with worldwide ones can productively build strength of worldwide highlights against different changes.
4. As worldwide highlights are not ready to adapt to geometric changes, these worldwide highlights can be proficiently joined with neighborhood highlights to reinforce them against both photometric and geometric changes.
5. Motion highlights are utilized to recognize recordings and however are not vigorous to content additions or different impediments which hinder the movement from being caught. Changes including revolution will alter the course of movement vectors and provide poor outcomes.
6. In expansion to vigor and separating capacities, the extricated highlight vector ought to be sufficiently smaller to perform a quick coordinating task, as reduced mark requires least storage room and performs likeness estimation in less calculation time.

4 Implementation

- The objective of undertaking to segment each reference video and inquiry video into short video portions
- A small video fragment is a gathering of comparable sequential edges
- It is not necessary that the fragments be a similar length and each section can choose one of its frames as a representative frame.

Steps for Novel Search Algorithm

For the most extreme movement relocations of h7, the proposed NS calculation uses an inside one-sided seek design with nine inspection focuses on a 5×5 window in the initial step rather than a 9×9 window in the 3SS. The focal point of the pursuit window is then moved to the point with least piece contortion measure. The pursuit window size of the following two stages relied upon the area of the base BDM (Block Distortion Measure) focuses. On the off chance that the base BDM point is found at the focal point of the inquiry window, the search will go to the last advance (Step 4) with 3×3 look window. Something else, the inquiry window measure is kept up in 5×5 for stage 2 or stage 3. In the last advance, the pursuit window is

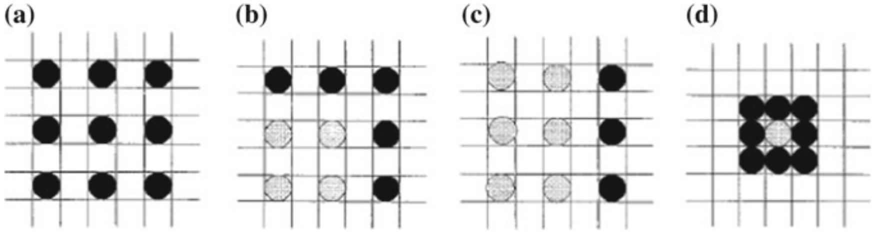


Fig. 1 **a** Nine-checking points pattern, **b** block distortion measure, **c** searching pattern strategy, **d** the Least BDM Point among these nine searching points

decreased to 3 × 3 and the inquiry stops at this little hunt window. The NS algorithm is summarized as follows

1. **Step 1**: A least BDM point is established from a nine-checking focuses design on a 5 × 5 window situated at the focal point of the 15 × 15 seeking region as appeared in Fig. 1a. In the event that the base BDM point is established at the focal point of the search window, go away to Step 4; generally go away to Step 2.
2. **Step 2**: The investigate pixel size is maintained in 5 × 5 to investigate and the outline will depend on the location of the previous least BDM point.

    (a) If the least BDM (block distortion measure) point is positioned at the area of the previously investigated pixel for five supplementary inspection points as shown in Fig. 2b are used.
    b) If the least BDM point is positioned at the center of horizontal or vertical axis of the earlier investigated pixel for three supplementary inspection points as shown in Fig. 2c are used. If the least BDM point is found at the middle of the investigate window, go away to Step 4; otherwise go away to Step 3.

3. **Step 3**: The searching model approach is equal as Step 2, but to finish it will go away to Step 4
4. **Step 4**: To investigate the pixel is reduced to 3 × 3 as shown in Fig. 2d and the direction of minimum BDM point among these nine searching points considered to the overall motion vector.

We can locate from the algorithm transitional strides of the NS might be left out and after that bounced to the last advance with a three by three pixel if whenever the base BDM point is situated at the focal point of the inquiry window. In light of this four-advance inquiry design, we can plaster the entire 12 × 12 uprooting window level just little pursuit windows, 4 × 4 and 3 × 3, are utilized. There are covered checking focuses on the 4 × 4 look skylight in the stage 2 and stage 3, along these lines the aggregate number of read-through focuses is fluctuated from (8 + 9) = 17 to (9 + 6 + 6 + 8) = 29. The most pessimistic scenario is the computational prerequisite of the NS is 29 square equivalents, it will occur for the evaluation of expansive development. Two cases of NS appeared in the various inquiry ways.

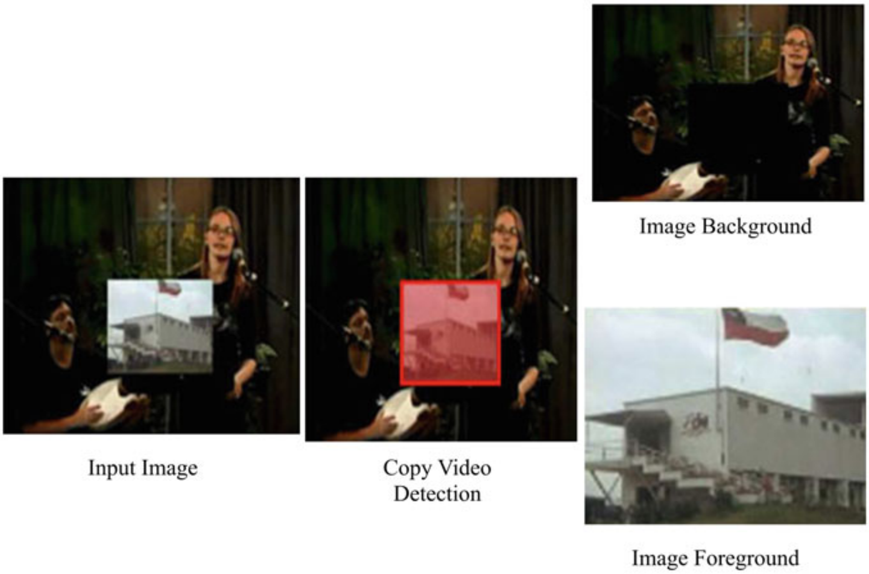


Fig. 2 Preprocessing task: A query video with PIP (Picture-in-Picture), detected and new query videos

For the high pursuit way, an aggregate of 25 inspection focuses is utilized with the assessed movement vector equivalent to (3, −8). For the most pessimistic scenario container of the lesser seek way, the quantity of checking focuses necessary is 29 and the assessed movement vector is equivalent to (−8, 8). We can locate that the computational multifaceted nature of NS (Novel Search) is just two square equal more than the Three Step Search while 6 piece coordinates not exactly the N3SS (New Three-Step Search) in the worst case.

5 Result Analysis

The objective of this preprocessing task is to divide every reference video and query video into short video segments. A small videotape segment is a collection of similar repeated frames and is not necessary that the segments be the matching length. Optionally, every segment can be choosing a single of its frames as a delegate frame. We separate a video into segments instead of attempt since dividing a video few segments and various corpora enclose duplicate shorter construction for division necessary in a single attempt. We just select keyframes for quality extraction and we will use an entire segment instead of just delegate frames. In our execution, we first separate the video into segments of fixed time-span, and then we join two repeated segments when all their frames are almost matching between them.

6 Conclusion

We propose the TASC (Transformation-Aware Soft Cascading) to sort out different multimodal locators in a falling and change mindful way, which is relied upon to accomplish high discovery precision while limiting the handling time. One productive usage is likewise created by using three ordinarily utilized multimodal highlights to develop four unique chains. A detection-on-copy-unit is TASC component is presented in which settles on the choice of duplicate location relying upon the closeness between their majority comparable CUs instead of the video-level comparability. To do as such, we likewise suggest a copy unit look calculation to discover a couple of copy units-based restriction calculation to locate the exact areas of duplicate sections with the attested CUs as the inside. To deal with the issue that the duplicates and non-duplicates are perhaps straightly indivisible in the component. We present an adaptable delicate choice limit technique in the TASC and after that suggest a bi-edge learning calculation for tough choice and use SVM classifier to delicate edge in view of the SIFT key point coordinating for delicate choice. Because of its incredible handling execution, the TASC is fit for fulfilling different prerequisites in down to earth duplicate discovery submission. The TASC-based framework will utilize this innovation in the video web index to distinguish the semantically and outwardly indistinguishable copies from video query items. For the future work, we plan to additionally advance the execution and adaptability of the TASC applications.

References

1. Liu H, Hong L, Xue X (2013) A segmentation and graph-based video sequence matching method for video copy detection. IEEE Trans Knowl Data Eng 25(8):1706–1718
2. Jiang M, Tian Y, Huang T (2012) Video copy detection using a soft cascade of multimodal features. In: Proceedings of the IEEE international conference on multimedia and expo (ICME'12), pp 374–379
3. Haitsma J, Kalke T (2012) A highly robust audio fingerprinting system. In: Proceedings of the international symposium on music information retrieval, pp 107–115
4. Tasdemir K, Cetin AE (2014) Content-based video copy detection based on motion vectors estimated using a lower frame rate. In: Proceedings of signal, image and video processing. Springer, Berlin, pp 1049–1057
5. Lei Y, Luo W, Wang Y, Huang J (2012) Video sequence matching based on the invariance of color correlation. IEEE Trans Circuits Syst Video Technol 22(9):1332–1343
6. Esmaeili MM, Fatourechi M, Ward RK (2011) A robust and fast video copy detection system using content-based fingerprinting. IEEE Trans Inf Forensics Secur 6(1):213–226
7. Barrios JM, Bustos B (2011) Competitive content-based video copy detection using global descriptors. Multimed Tools Appl https://doi.org/10.1007/s11042-011-0915-x (Springer Science+Business Media)
8. Song J, Yang Y, Huang Z, Shen HT, Hong R (2013) Multiple feature hashing for large scale near-duplicate video retrieval. IEEE Trans Multimedia 15(8):1997–2008

Color Image Encryption: A New Public Key Cryptosystem Based on Polynomial Equation

P. K. Kavitha and P. Vidhya Saraswathi

Abstract Image security is the need of today's era. The security task is being convoluted as there are a variety of data to secure as different levels of data security is needed based on the data type. Encryption is the maximum expedient approach to guarantee the safety of images over public networks. We have proposed a new public key cryptosystem based on polynomial equation. This approach has been desired because of their resistance to the cryptanalysis attacks. Parameters such as correlation coefficient, entropy, histogram analysis are used for the efficiency of our proposed method. The theoretic and simulation result provides a high quantity of protection. Thus, it is suitable for practical use in secure communication.

1 Introduction

Due to the enormous growth of network era, numerous facts can be transmitted and posted fluently and quick. To store the resources, hundreds of digital images of some particular application fields additionally had been transmitted through the public network. There can be plenty of secret and sensitive records contained in the images [1]. So, we have to encrypt those digital images earlier before transmission through the open network. The protection of digital images has been a focal point and one of the important significant branches in the information protection area. Because of its unique features such as lively and intuitionist, image has been one of the favored forms expressed through network. An image has high redundancy and strong correlation between pixels. So, tough security is needed for storage and transmission of images through networks. Most of this information is collected and saved on electronic computers and transferred all over the network [1]. If eavesdroppers

P. K. Kavitha (✉) · P. Vidhya Saraswathi
Department of Computer Applications, Kalasalingam Academy of Research and Education, Krishnankoil 626126, Tamil Nadu, India
e-mail: pkkavitha78@gmail.com

P. Vidhya Saraswathi
e-mail: vidhyasaraswathi.p@gmail.com

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_8

found the secret images regarding enemy positions then security could lead to declination of clash. Protecting images is an ethical and authorized requirement. A lot of images are usually used in one kind of outstanding procedure. Image encryption is a vital function within the field of data hiding. Image encryption method prearranged an image by applying some mathematical formulas, so no one gets the original image [2]. In this paper, color image encryption is done using polynomial equation. The cryptography process requires some algorithms for encrypting the data. Two categories of cryptography are symmetric and asymmetric [2]. In symmetric key cryptography, both the sender and the recipient use the similar secret key and both of them know the key. In sender side, the data is encrypted using the secret key and the recipient decrypts the information using the same secret key. Key depends on the nature of the key. Examples of symmetric key methods are data encryption standard (DES), advanced encryption standard (AES), international data encryption algorithm (IDEA). In asymmetric key cryptography, a couple of keys used are public and private keys. It has six substances. There are plain texts, encryption algorithm, public key, private key, cipher text, and decryption algorithm. Examples of asymmetric key methods are Diffie–Hellman Key Exchange [3], RSA Encryption Algorithm, Elliptic Curve Cryptography (ECC), and ElGamal Cryptosystem [2]. In this work, we follow an innovative public key cryptosystem corresponding to polynomial equation for more security of color images.

The rest of this paper is structured as follows. Section 2 describes the related work. Section 3, in brief, explains the concept of polynomials that is the core of this report. Section 4 describes the results and discussion of the proposed scheme. Section 5 produces experimental results and evaluation parameters to validate the proposed color image encryption algorithm. Section 6 concludes the paper.

2 Related Work

Seyedzade et al. [4] presented a precise set of rules for image encryption primarily based on SHA-512. The crucial idea of the set of regulations is to use one half of image data for encryption of the alternative half of the image equally. Individual characteristics of the set of rules are high protection, immoderate sensitivity and excessive pace that may be applied for encryption of grayscale and color images. The goal is to increase the image entropy. Hashim et al. [5] selected a public key cryptosystem known as the ElGamal Cryptosystem. The ElGamal cryptosystem over a primitive root of a huge excessive is utilized in messages encryption inside the unfastened privacy guard software application, present day-to-day variations of pretty good privacy, and other cryptosystems. Here, change of the cryptosystem is via making use of it over gray and color images. That is probably by using remodeling an image into its corresponding matrix using MATLAB program. The encryption and decryption algorithms is applied over it. This new modification should make the cryptosystem greater immune toward a few destiny assaults considering that breaking this cryptosystem relies upon on fixing the discrete logarithm hassle and it is actually now not feasible with massive pinnacle numbers. Khachatrian et al. [6] added a public key encryption and virtual signature system based on permutation polynomials is evolved. It is proven that for the new device to obtain a comparable safety with conventional public key structures based totally on both Discrete logarithm or Integer factorization issues, significantly a lot of processing period is wanted to ensue in a large acceleration of public key operations. Sokouti et al. [7] carried out a new GGH encryption algorithm, where GGH is a new public key cryptosystem that is corresponding to the closest vector problem. It encrypts records in the matrix form and makes it appropriate for image encryption. Tentative research converting one pixel in GGH encryption does not have enough effect at the whole encryption output, so a new method is used which uploads padding to the original image in advance than making use of GGH encryption method. Then the forward and backward snail tour XORing into the original image is applied. Based on the very last consequences which correspond to the proposed approach, converting one pixel affects the entire image and additionally influences the cipher image pretty right this is encrypted with the resource of GGH set of rules. Dwivedi et al. [8] have been planned to offer at ease cryptosystem to transmit images. A new curved scrambling method expanded the complexity of key generation method, consequently difficult to decode via the hacker. Also diffusion technique adjusts the corresponding pixel values and hence fend off repeated permutation. All these strategies assist to layout a greater complexity and comfy cryptosystem to transmit the data over the insecure community. Wagh et al. [9] offered a unique image encryption set of rules based on permutation polynomials over integer rings which makes a strive to conquer the restrictions of current strategies. Here, the authentic image is scrambled by way of applying permutation polynomial to its rows and columns. The estimated changed price in a wide variety of pixels and Unified Averaging Changing Intensity measures are near to theoretical values and are comparable to present encryption strategies.

3 Proposed Methodology

The proposed method is a new public key cryptosystem based on polynomial perspective. Original image with public key image produces an encrypted image. In the recipient side, to get the authentic image receiver's own private key image is used. The experimental effects proved that this new cryptosystem is further competent when compared with common existing techniques used for image encryption.

3.1 Polynomial

A polynomial of degree n inside variable x is a function defined by means of

$$f(x) = k_0 + k_1 x_1 + k_2 x_2 + \cdots + k_{n-1} x_{n-1} + k_n x_n$$

Here, n is the degree of the polynomial k_n = the leading coefficient, k_0 is the constant term, k is real numbers and also the coefficients of the polynomial.

3.2 Encryption Phase

In this section, we discuss the algorithm design and implementation of the proposed image encryption and decryption method using polynomial. Row and column of an image is depicted as r, c. The row and column value is mentioned as $r = c = 256$. The decryption scheme is in direct contrary to an encryption procedure. The proposed approach is based on public key cryptosystem and for this reason, encryption and decryption require a pair of keys.

Algorithm for Encryption

1. Input: (original image, public key image, private key image)
2. Compute: Mean value of private key image
3. Unimodular matrix (private key image, mean value)
4. If pixel value of private key image > mean value
5. Set the unimodular matrix value as 1
6. Else
7. Set the unimodular matrix value as 0
8. End
9. Use the polynomial equation to choose the pixel in the unimodular matrix
10. For r = 1: (size of the original image, 1)
11. For c = 1: (size of the original image, 2)
12. Obtain (first pixel, second pixel) from original image
13. Choose the same index as (first pixel, second pixel) from public key image
14. Perform XOR between original image and public key image
15. Goto next pixel based on polynomial graph
16. Repeat steps 10–14 until all of the plain image pixels are processed
17. End
18. End
19. Output: encrypted image.

4 Result Analysis

The planned polynomial-based approach has been developed in MATLAB R2013a software. Color image is taken as an input image. It has been resized into 256 × 256 pixels. Pick another image as a public key image and resized into 256 × 256 pixels. Then, the private key image could be read and resized into 256 × 256 pixels. Then these three images changed into grayscale images. Computed the mean value

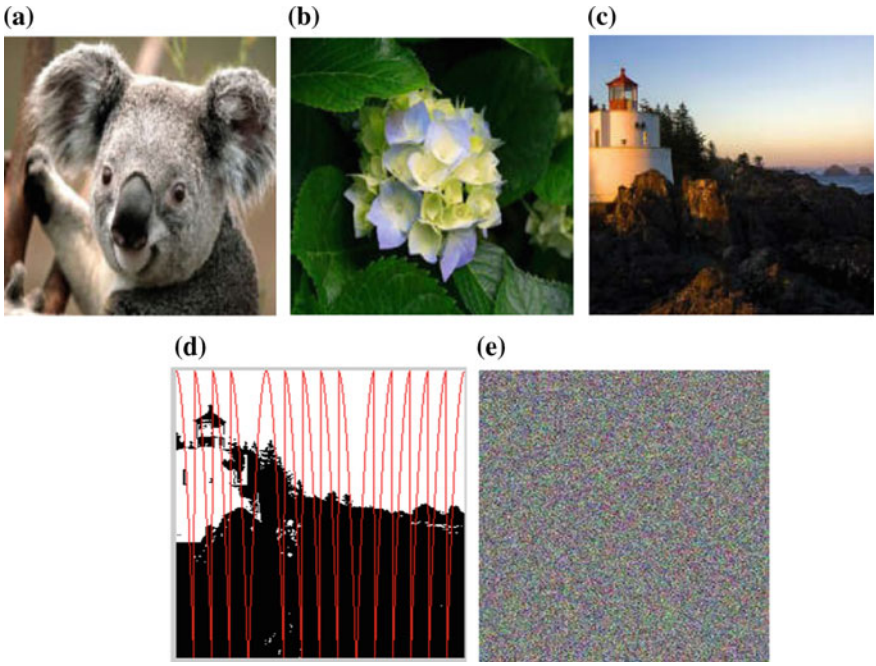


Fig. 1 Color images **a** Original image. **b** Public key image. **c** Private key image. **d** Polynomial in unimodular matrix. **e** Encrypted image

of a private key image. Based on this mean value, generate the unimodular matrix (i.e., 0s and 1s). Suppose MV is the mean value of the private key image. UM is the unimodular matrix. $[X, Y]$ is the pixel value of unimodular matrix.

$$\text{If } PR > \text{MV then}[X, Y] = 1, \text{ else } [X, Y] = 0.$$

That is the value set as 1, if the mean value is greater than private key image value and the value set as 0, if it is less than the pixel value of the private key image. Then the unimodular matrix is created. Now, the new polynomial-based approach can be almost carried out to an original image. Figure 1 demonstrate the step by step encryption of color image using our proposed approach.

In the encryption phase, the public key image is used. From the public key image, take an index value from the polynomial graph in a unimodular matrix image. The same index is taken from the original image and calculate the value of encrypted image by using the original image pixel value and the public key pixel value. Then an encrypted image is produced by XORing the two values from the original image and public key image.

Private key is used to decrypt an encrypted image. On behalf of decryption, the recipient receives an encrypted image and use the contrary approach of the private key

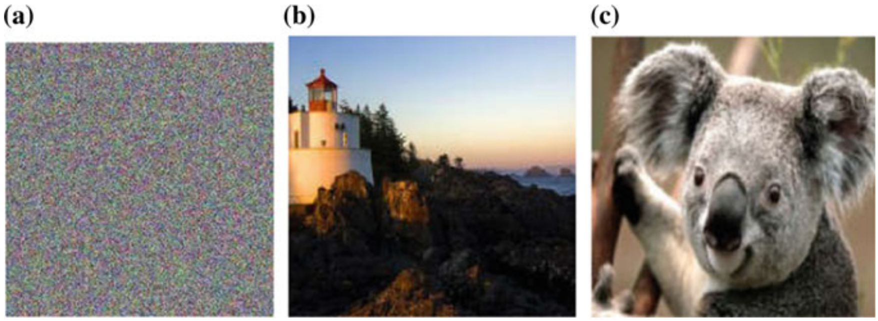


Fig. 2 Decryption process **a** Encrypted image. **b** Private key image. **c** Decrypted image

image to an encrypted image to get the original image. Here, XORing two values are obtained from the encrypted image and the private key image depends on the polynomial graph in unimodular matrix. The decrypted image is as shown in Fig. 2c.

5 Performance Measurement

The proposed polynomial-based encryption approach is tested on MATLAB R2013a software. Proposed method should defy along with recognized attacks like cipher-text attack, known-plain text, brute-force attack, and differential attack. Safety measures analysis plays on the proposed approach simultaneously with the statistical analysis and differential analysis as follows.

5.1 Histogram Analysis

Histogram graph depicts the wide variety of pixels in an image at dissimilar intensity values observed in an image. The histogram will show 256 numbers displaying the distribution of pixels among the one's of the grayscale values. Proposed encryption approach's competence is checked with the aid of the distribution of grayscales in an encrypted image.

It needs to be uniform and notably distinctive from histograms of an Original Image. From Fig. 3a–d are the histograms of the authentic input image and an encrypted image. Histogram of an encrypted image is identical and widely contradictory from an original image. It doesn't give any evidence to utilize any statistical attack on the proposed polynomial-based image encryption approach.

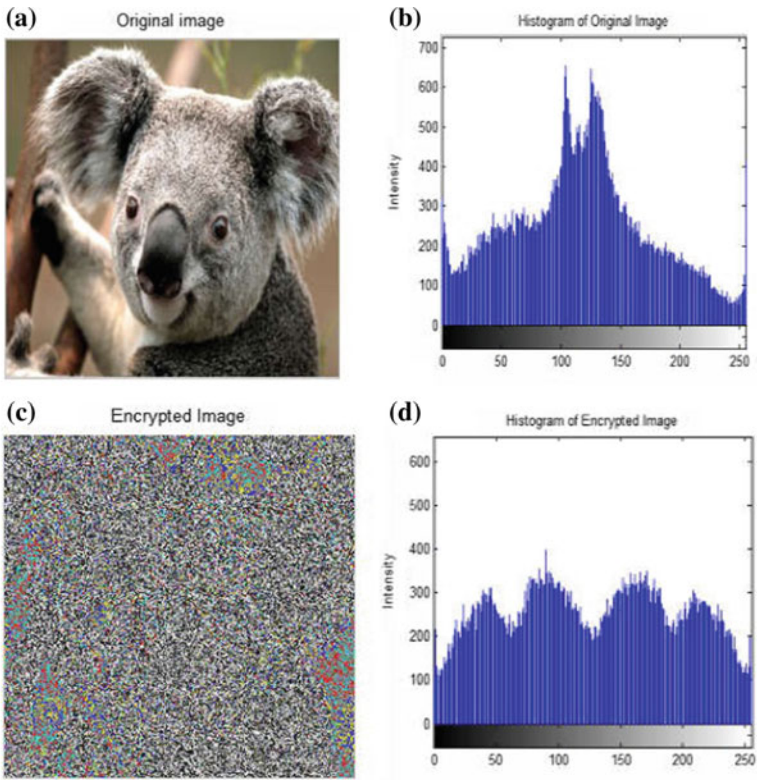


Fig. 3 **a** Input image (Koala). **b** Histogram of input image. **c** Encrypted image of Koala. **d** Histogram of encrypted image

Table 1 Mean values of original and encrypted image for RGB channels

Test image: Koala	Original	Encrypted
Red mean	125.40	127.50
Green mean	114.74	127.40
Blue mean	101.97	127.35

5.2 RGB

From the outcome, the original image and encrypted image histograms and their equal histograms for Red, Green, and Blue are as shown in Fig. 4. The encrypted image is again and again spread and hugely unlike from the histograms of an input Image. Mean values of original and encrypted image for RGB channels displayed in Table 1.

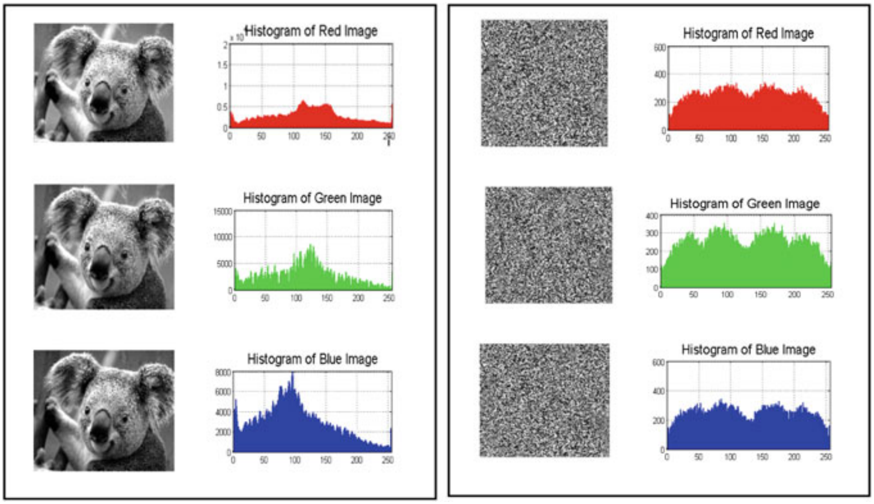


Fig. 4 RGB histograms for original and encrypted Koala image

5.3 Correlation Analysis

Correlation is a quantity of the link between two variables or pixels in an image. Adjacent pixel correlation means, if two pixels from an image taken are the nearest pixels in an image, then say that these two pixels are strongly related. If the two pixels are distant, then they are not as much of correlated. Correlation among the adjacent pixels value can be very high. The competent approach can diminish the relationship between the pixels. First, prefer 2000 pixel pairs from an input image. Figure 4 suggests the connection part of adjacent pixels in the input image and cipher image (koala image of size 256 × 256) for the personalized encrypted image. Conversely, the pair of adjacent pixels in an input image is extremely interrelated. Correlation coefficient of all pair can be calculated using the formula:

$$\textbf{Cor} = \frac{k\left(\sum(m,n) - \sum m \sum n\right)}{\sqrt{\left[k\left(\sum(m)^2 - \left(\sum m\right)^2\right)\right]\left[k\left(\sum(n)^2 - \left(\sum n\right)^2\right)\right]}}$$

Here,

Cor Correlation Coefficient
k The number of pairs of data
$\sum mn$ Range of the products of paired data
$\sum m$ Sum of the m statistics
$\sum n$ Sum of n data
$\sum m^2$ Sum of Squared m

Table 2 Correlation coefficient and entropy value

Test image: Koala	Input image	Encrypted image
Horizontal	0.9460	−0.0071
Vertical	0.9568	−0.0058
Diagonal	0.9240	0.0041
Entropy	7.8292	7.9539

$\sum \boldsymbol{n}^2$ Sum of Squared ***n***

Correlation coefficients are 0.9460 and −0.0071, respectively, for input image and encrypted image. These two values are distant from each other. Related outcome for vertical and diagonal directions was attained like in Table 2.

5.4 Entropy Value

It shows the quality of an image. Entropy is expressed in terms of bits. Entropy value of an input image (Koala) and encrypted image is like in Table 2. Use the following formula for entropy value, we have:

$$E = \sum_{i=0}^{k-1} g(y)\log_2(g(y))$$

Here,

E = Entropy Value, k = Input image gray value from 0 to 0.255 and $g(y)$ = Probability of the amount of symbol y

6 Conclusion

A new polynomial-based public key cryptosystem is proposed in this paper and implemented in MATLAB R2013a programming software. The conclusion can be made from the results and discussions are done. This proposed polynomial-based color image encryption algorithm gives more security compared to the existing methods. This paper offers a solution to fulfill the increasing requirements for secure image transmission through networks. Polynomial-based cryptosystem supports encryption of various image sizes. Experimental analysis shows that our new polynomial-based image encryption technique is efficient.

References

1. Gonzalez W (2008) Digital image processing, 2nd edn. Prentice Hall, Englewood Cliffs
2. Stallings W (2014) Cryptography and network security, principle and practice, 6th edn
3. Tambe Chaitali A, Gadilkar Rupali S, Pawar Vimal T (2017) A survey on new image encryption algorithm based on Diffie-Hellman and singular value decomposition. Int J Innov Res Comput Commun Eng
4. Seyedzade SM, Atani RE, Mirzakuchaki S (2010) A novel image encryption algorithm based on hash function. In: 6th Iranian conference on machine vision and image processing, 27–28 Oct 2010, pp 1–6
5. Hashim HR, Neamaa IA (2014) Image encryption and decryption in a modification of ElGamal cryptosystem in MATLAB. Int J Sci: Basic Appl Res 14(2):141–147
6. Khachatrian G, Kyureghyan M (2014) A new public key encryption system based on permutation polynomials. In: IEEE international conference on cloud engineering
7. Sokouti M, Zakerolhosseini A, Sokouti B (2016) Medical Image encryption: an application for improved padding based GGH encryption algorithm. Open Med Inf J
8. Dwivedi N, Gupta RK, Agarwal S (2017) Image encryption using curved scrambling and diffusion. Int J Eng Technol
9. Wagh NB, Kolhekar M (2017) A novel approach utilizing permutation polynomials over integer rings as a cryptological application for effective encryption of digital images., IJCSE 5(1). E-ISSN 2347–2693

Thermal Imaging of Abdomen in Evaluation of Obesity: A Comparison with Body Composition Analyzer—A Preliminary Study

S. Sangamithirai, U. Snekhalatha, R. Sanjeena and Lipika Sai Usha Alla

Abstract Changes in body composition parameters lead to obesity progression which has an impact on body metabolism. The aim and objectives of the study was: (i) to estimate body composition parameters such as fat-free mass, skeletal muscle mass, lean body mass, subcutaneous fat mass using body composition analyzer; (ii) and to measure mean surface temperature of various regions of the body such as abdomen, chest, forearm front and back, shank front and back using thermal imaging method. Adult volunteers of age 20–29 years were participated in this preliminary study. The body composition parameter such as fat mass, fat free mass, muscle mass, and skeletal muscle mass was measured using body composition analyzer. The thermal imaging of the forearm, abdomen, and shank region was obtained and the skin surface temperature was measured in both normal and obese. The FFM negatively correlated with mean surface temperature at the abdomen region and found to be highly significant with $p < 0.05$. In this study, BCA parameters correlated significantly with the skin surface temperature measured at forearm and abdomen region in obese subjects.

1 Introduction

Obesity refers to abnormal fat accumulation in the body, which leads to impairment to the health [1]. According to the World Health Organization (WHO) criteria, the

S. Sangamithirai · U. Snekhalatha (✉) · R. Sanjeena · L. S. U. Alla
Biomedical Engineering Department, SRMIST, Kattankulathur, Tamil Nadu, India
e-mail: sneha_samuma@yahoo.co.in

S. Sangamithirai
e-mail: sangamithirai.s.kumar@gmail.com

R. Sanjeena
e-mail: sanjeenar.2297@gmail.com

L. S. U. Alla
e-mail: sophiadumbledore@yahoo.com

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_9

simple tool to evaluate the obesity is through calculating the body mass index (BMI). The BMI >30 kg/m^2 is considered as obese as per WHO criteria. The major risk factors of obesity include diabetes, cardiovascular disorder (CVD) and cancer, sleep apnea, and depression. The prevalence of obesity worldwide is 13% in the year 2016 [2]. By the year 2030, the increase in obesity level is expected to be 47%, 39%, and 35% in the US, Mexico, and England, respectively [3]. The prevalence of obesity in India is 9.3% in the male population and 12.6% in the female population [4].

The different diagnostic methods for the obesity assessments are as follows: BMI, waist circumference, skinfold thickness, waist to hip ratio, bioelectric impedance method, dual energy X-ray absorptiometry, computerized tomography (CT) and magnetic resonance imaging (MRI). Body composition parameters estimate the fat mass, fat free mass, muscle mass and segmental muscle mass from the subjects using body composition analyzer. Several Researchers have performed the body composition analysis in obese subjects using bioimpedance analysis method and estimated the fat mass, fat free mass, BMI, BMR, waist circumference, hip circumference, waist–hip ratio, etc. [5–8].

Thermal imaging is a noninvasive, noncontact imaging modality, which works on the principle of our body emits radiation which is sensed by infrared cameras. The thermal imaging method is used in a variety of applications in the medical field such as in assessment of breast cancer [9], diabetic mellitus [10], rheumatoid arthritis [11], cardiovascular disorders [12], temporomandibular joint disorders [13], vascular disorders [14], ocular [15] and skin diseases [16], etc. Infrared thermography is used as a research tool for assessment of obesity in various regions of the body.

The body surface temperature gets reduced in obese people due to loss of metabolic heat, because of core body temperature stimulates the vasoconstriction which is high in obesity [17]. Savastano et al. conducted the study on obesity to study the relationship between the adiposity and hand and abdomen temperature in both normal and obese adults [17]. Chudeka et al. evaluated the changes in body composition parameters and body surface temperature measured using thermal imaging in females after abdominal liposuction [18]. Chudeka et al. investigated the level of visceral and subcutaneous adipose tissue which affects the core body surface temperature and performed the thermal mapping of skin surface temperature in obese women [19]. Jalil et al. compared the body temperature of the abdomen, neck, and hand using thermal imaging in obese and normal subjects during oral glucose tolerance test [20].

The aim and objectives of the study was: (i) to estimate body composition parameters such as fat free mass, skeletal muscle mass, lean body mass, subcutaneous fat mass using body composition analyzer; (ii) and to measure mean surface temperature of various regions of body such as abdomen, chest, forearm front and back, shank front and back using thermal imaging method.

2 Methodology

2.1 Participants

Volunteer adults (age: 20–29 years) participated in this study. It was explained that the study investigated the relation between body composition parameters and thermal images in adults. None of the participants were undergoing treatment related to obesity or were using medication that affected heat balance in the body. The consent of all participants were obtained.

2.2 Experimental Design

Each volunteer visited the Medical Instrumentation Laboratory in the Department of Biomedical Engineering, SRMIST, Kattankulathur, India. First, height and weight were measured and body composition parameters such as fat mass, fat free mass, muscle mass, skeletal muscle mass were analyzed. Second, the participants were told to place their forearm on a table of height of about 2 feet. A thermal camera was set at a focal length of 600 mm from the table. Regional body temperatures were assessed in the forearm, shank and abdomen areas. The room was temperature controlled (21 °C) and usual comfortable clothing was worn by the subjects.

2.3 Regional Body Temperature Measurements

The thermal camera used a 640 * 512 array of sensors sensitive to changes in temperature. For further off line analysis, the thermal images acquired were stored on a computer. Participants were advised to stay still during imaging, to keep motion artifacts to a minimum.

During forearm imaging, subjects were seated on a chair with their arm placed on a black background to minimize drafts. Infrared images were acquired in each region for 10 min duration. During abdominal imaging, a curtain was drawn to lessen drafts while the participants lay in a supine position on the table. Their clothing was adjusted in such a way as to allow complete abdominal exposure from the level of the xiphoid process to 2.24 cm above the symphysis pubis. A 5 min acclimatization period was set with the patient at rest, after which, sequential thermographic images were acquired for 5 min with his or her abdomen exposed. The shank area was imaged after asking the participants to sit on a table 2 feet above the floor. Lens was focused on the exposed region and images were acquired for 5 min.

Statistical Package for the Social Sciences (SPSS) software was used to process the infrared images collected. Temperature values were extracted from each region of interest (ROI) for the sequential images. The subcutaneous fat deposit in the abdomen was selected depending on the most consistent region among the participants. A rectangular ROI (100 pixels total) 1.2 cm subadjacent to the umbilicus was analyzed.

2.4 Body Composition Measurements

Measurement of body composition in humans acts as a response to conditions such as obesity, where body fat and BMD allow for clinical diagnosis. The body composition analyzer is a single frequency device using single-point load weighing system in the scale platform for different segments of the body such as right arm, left arm, right leg, and left leg. The measurements for fat, total body water (TBW), body mass index (BMI), fat free mass (FFM), basal metabolic rate (BMR), visceral fat rate (VFR), extracellular water (ECW), muscle mass, bone mass, and fat mass were recorded.

3 Results and Discussion

The demographic variables and BCA parameters such as muscle mass and fat mass at hand and leg for normal, overweight, and obese subjects were given in Table 1. The basic demographic variables such as BMI, FFM, and BMR are increased by 50%, 12.5%, and 10.2%, respectively, in obese subjects compared to normal subjects. The muscle mass at right arm shows the highest significant difference of 17.76% between normal and obese subjects compared to other regions such as left arm, left leg and right leg. Similarly, the fat mass at right arm depicts the greater difference 62.3% in obese compared to normal subjects in other regions.

The mean skin surface temperature measured at the forearm, abdomen, and shank region for normal, overweight, and obese were listed in Table 2. The decreased temperature of about −2.78% exhibits in the obese subject compared to normal in the forearm region. Similarly, the lesser temperature was observed in obese subjects compared to normal in abdomen and shank region. Among the three regions studied, abdomen shows the highest temperature difference between the normal and obese subjects.

The correlation between the BCA parameters and thermal imaging parameters in obese subjects were depicted in Table 3. The FFM negatively correlated with mean surface temperature at the abdomen region and found to be highly significant with $p < 0.05$. Similarly the other BCA parameters such as BMR, Muscle mass, bone mass, and fat mass significantly produced a negative correlation with the mean surface temperature. In the case of forearm region, FFM, BMR, muscle mass, bone mass, and fat mass correlated with average temperature at the significance level $p < 0.01$. But the BCA parameters had not correlated significantly with the mean surface temperature at the shank region.

Table 1 Body composition parameters for normal, overweight, and obese subjects

BCA	Baseline variables			% Difference
		Normal	Obese	
Parameters	AGE	24 ± 6.9	23.62 ± 6.45	−1.608
	BMI	20.78 ± 2.39	41.5 ± 4.2	50
	FFM	36.05 ± 5.71	41.22 ± 7.8	12.5
	BMR	1126.25 ± 1.77	1255 ± 207.8	10.2
Muscle mass	LA	1.53 ± 0.3	1.83 ± 0.42	16.4
	RA	1.583 ± 0.4	1.925 ± 0.43	17.76
	LL	6.608 ± 0.9	7.55 ± 1.925	12.47
	RL	6.75 ± 0.98	7.725 ± 1.6	14.45
Fat mass	LA	0.59 ± 0.28	1.46 ± 0.4	60
	RA	0.533 ± 0.27	1.412 ± 0.4	62.3
	LL	3.15 ± 0.96	5.837 ± 1.3	46
	RL	3.2 ± 1.0027	5.787 ± 1.22	44.7

Table 2 Mean skin surface temperature measured at the forearm, abdomen, and shank region for normal, overweight and obese

Region of interest	Normal ($N = 11$)	Obese ($N = 10$)	% Difference
Forearm	33.21 ± 1.6	32.31 ± 1.81	−2.78
Abdomen	33.17 ± 1.03	32.86 ± 1.16	0.94
Shank	32.65 ± 1.371	32.61 ± 1.396	0.12

Table 3 Correlation between the BCA and temperature parameters at forearm, abdomen, and shank region

BCA parameters	Forearm	Abdomen	Shank
FFM	−0.104	−0.058	−0.078
BMI	−0.037	−0.13	−0.028
FFM	−0.337*	−0.578**	−0.208
BMR	−0.3805*	−0.592**	−0.205
Muscle mass	−0.3045*	−0.599**	−0.164
Bone mass	−0.416*	−0.614**	−0.231
Fat mass	−0.381*	−0.223	−0.115

*$p<0.05$
**$p<0.01$

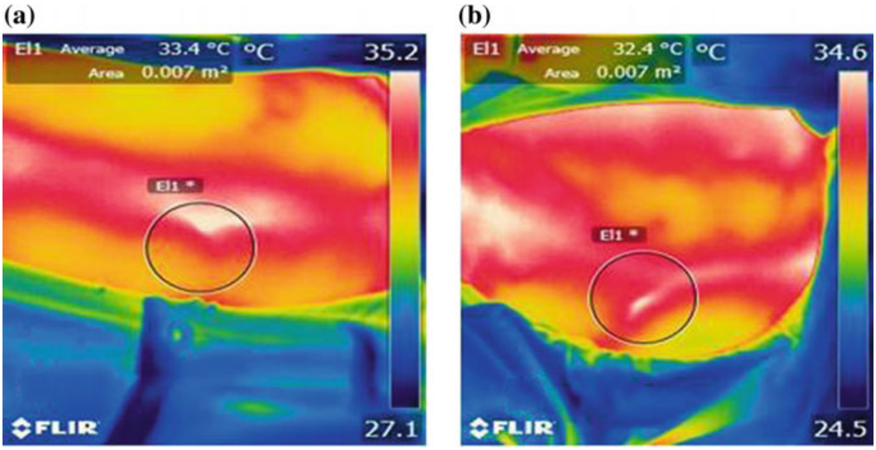


Fig. 1 **a** and **b** represent the forearm of normal and obese, respectively

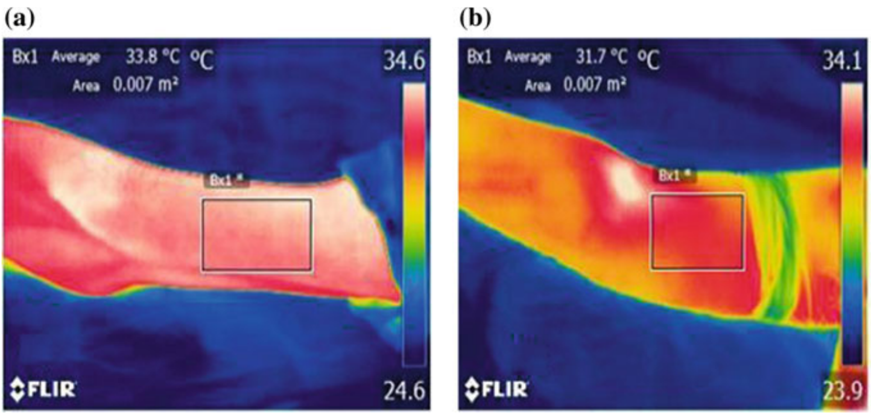


Fig. 2 **a** and **b** indicate abdomen of normal and obese, respectively

Figure 1a represents temperature measured at forearm by keeping circular ROI as fixed. The average temperature obtained at forearm region for the normal subject was 31.7 °C, whereas the measured temperature for the obese subject at forearm region depicts 31.4 °C as mentioned in Fig. 1b. Figure 2a indicates the temperature measured at abdomen region for normal subject in which the average temperature displays 32.6 °C. Figure 2b illustrates the average temperature as 31.7 °C in abdomen region for obese subject. Figure 3a depicts the average temperature as 31.3 °C measured at shank region for normal subject, whereas for obese subject the average temperature measured at shank region was 30 °C.

Savastono et al. obtained the mean surface temperature measured at abdomen region for normal and obese subjects as 33 ± 0.3 °C and 31.8 ± 0.3 °C, respectively [18]. They found the negative correlation ($\gamma = -0.45$, $p < 0.005$) between the fat

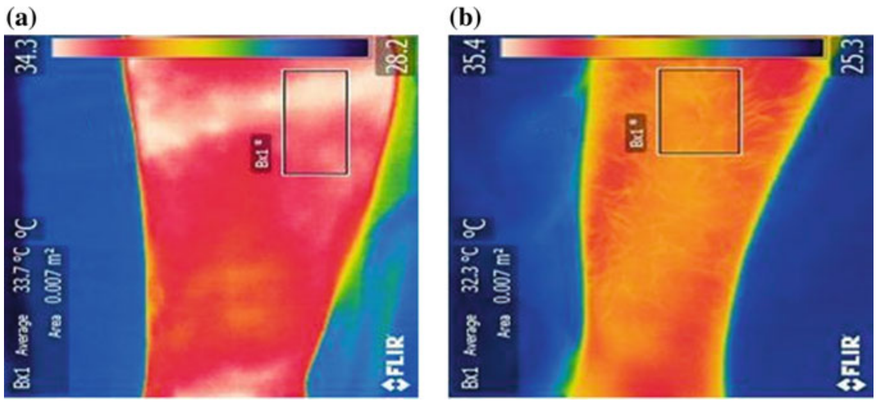


Fig. 3 **a** and **b** depict the temperature measured in the shank region of normal and obese, respectively

mass and abdominal temperature. Also, they found the significant difference between the two groups such as normal and obese in the analysis of BCA parameters such as fat mass and fat free mass. Chudecka et al. correlated the BCA parameters such as visceral fat mass, subcutaneous fat mass, and skeletal muscle mass with the abdomen temperature measured by infrared imaging [21]. They obtained the significant negative correlation between the visceral fat mass and abdominal temperature as $r = -0.7$, $p = 0.001$, and between the subcutaneous fat mass and abdominal temperature as $r = -0.62$, $p = 0.006$. In our study, significant negative correlation achieved between the fat free mass and abdominal temperature ($r = -0.57$, $p < 0.01$), and between muscle mass and abdominal temperature ($r = 0.59$, $p < 0.01$). This negative correlation is due to as BCA parameters increases in obese subjects, the temperature measures at abdomen region decreases. The reason behind was heat dissipation occurs at adipose tissues in nearby fat deposition regions in the abdomen.

4 Conclusion

In this study, BCA parameters correlated significantly with the skin surface temperature measured at forearm and abdomen region in obese subjects. Among the three regions studied, abdomen region shows greater temperature difference between the normal and obese subjects. The muscle mass and fat mass measured at forearm region exhibit higher value in obese subjects compared to normal. The BMI, BMR, FFM increased in obese subjects compared to normal and produces a significant difference. Hence in young age population, obesity exists due to sedentary lifestyle and leads to increased risk of causing other complications such as diabetes, CVD, etc. Hence, dietary control and maintaining the metabolic rate prevents obesity in young adults.

Acknowledgements The authors would like to express their sincere gratitude to Head, Biomedical Engineering Department SRMIST for providing the infrastructure and equipment facility.

References

1. Ofei F (2005) Obesity-a preventable disease. Ghana Med J 39:98–101
2. http://www.who.int/mediacentre/factsheets/fs311/en/
3. https://www.oecd.org/els/health-systems/Obesity-Update-2017.pdf
4. Nagendra K, Nandini C, Belur M (2017) A community based study on prevalence of obesity among urban population of Shivamogga, Karnataka, India. Int J Community Med Public Health 4:96–99
5. Shishkova A, Petrova P, Tonev A, Bahlova P, Softov O, Kalchev E (2007) Analysis of body composition in overweight and obese women using bioimpedance (BIA) system. J IMAB 13:8–12
6. Ayvaz DNC, Kilinc FN, Pac FA, Cakal E (2011) Anthropometric measurements and body composition analysis of obese adolescents with and without metabolic syndrome. Turk J Med Sci 41:267–274
7. Namwongprom S, Rerkasem K, Wongthanee A, Pruenglampoo S, Mangklabruks A (2014) Relationship between body composition parameters and metabolic syndrome in young Thai adults. J Clin Res Pediatr Endocrinol 6:227–232
8. Hunma S, Ramuth H, Miles-Chan JL, Schutz Y, Montani JP, Joonas N, Dulloo AG (2016) Body compostion-derived BMI cut-offs for overweight and obesity in Indians and Creoles of Mauritius: comparison with Caucasians. Int J Obes 40:1906–1914
9. Han F, Shi G, Liang C, Wang L, Li K (2015) A Simple and efficient method for breast cancer diagnosis based on infrared thermal imaging. Cell Biochem Biophys 71:491–498
10. Sivanandam S, Anburajan M, Venkatraman B, Menaka M, Sharath D (2012) Medical thermography: a diagnostic approach for type 2 diabetes based on non-contact infrared thermal imaging. Endocrine 42:343–351
11. Snekhalatha U, Rajalakshmi T, Gopikrishnan M, Gupta N (2017) Computer-based automated analysis of X-ray and thermal imaging of knee region in evaluation of rheumatoid arthritis. Proc Inst Mech Eng, Part H: J Eng Med 231:1178–1187
12. Jayanthi T, Anburajan M, Menaka M, Venkatraman B (2014) Potential of thermal imaging as a tool for prediction of cardiovascular disease. J Med Phys 39:98–105
13. Clemente M, Coimbra D, Silva A, Aquiar Branco C, Pinho JC (2015) Application of infrared thermal imaging in a violinist with Temperomandibular disorder. Med Probl Perform Art 30:251–254
14. Bagavathiappan S, Saravanan T, Philip J, Jayakumar T, Raj B, Karunanithi R, Panicker TMR, Paul Korath M, Jagadeesan K (2009) Infrared thermal imaging for detection of peripheral vascular disorders. J Med Phys 34:43–47
15. Kawali AA (2013) Thermography in ocular inflammation. Indian J Radiol Imaging 23:281–283
16. Gurjarpadhye AA, Parekh MB, Dubnika A, Rajadas J, Inayathullah M (2015) Infrared imaging tools for diagnostic applications in dermatology. SM J Clin Med Imaging 1:1–5
17. Jongh D, Serne RT, Ijzerman RG, De vries RG, Stehouwer C (2004) Impaired microvascular function in obesity: implications for obesity-associated micro-angiopathy hypertension and insulin resistance. Circulation 109:2529–2535
18. Savastano DM, Gorbach AM, Eden HS, Brady SM, Reynold JC, Yanovski JA (2009) Adiposity and human regional body temperature. Am J Clin Nutr 90:1124–1131
19. Chudecka M, Dmytrzak A, Lubkowska A (2016) Changes in selected Morphological parameters and body composition as well as Mean body surface temperature assessed by thermal imaging in women after abdominal liposuction. Central Euro J Short Sci Med 14:21–26

20. Jalila B, Hartwigb V, Moronia D, Salvettia O, Benassia A, Jalile Z, Guiduccib L, Pistoiac L, Tegrimic TM, Galvanc AQ, Iervasib G, Abbateb AL (2017) Near Infrared and thermal imaging of normal and obese women during oral glucose tolerance test (OGTT). In: 14th International workshop on advanced infrared technology and applications, Quebec city, Canada, pp 144–148
21. Chudecka M, Lubkowska A, Podhorodecka AK (2014) Body surface temperature distribution in relation to body composition in obese women. J Therm Biol 43:1–6

A Review on Protein Structure Classification

N. Sajithra, D. Ramyachitra and P. Manikandan

Abstract A massive amount of sequence data is gradually produced by the genome projects that have to be annotated in terms of structure, molecular, and biological functions. In structural genomics, the aim is to resolve several protein structures in an efficient way and to exploit the solved protein structures for assigning the biological function to theoretically solved protein structures. In earlier stages, the protein structures are classified manually in a successful manner and now it suffers from updating problem because of the high throughput of recently solved protein structures. To overcome this issue, several data mining techniques have been examined for the structural classification of the protein world. This review article presents an overview of the existing classification techniques, databases, tools, and performance metrics used for evaluating the performance of protein structure classification algorithms.

1 Introduction

It is particularly a fact for the proteins, whose functions are honestly related to their 3D structures [1, 2]. Protein structure classification is used to identifying the hierarchal clusters of proteins with related structures and it plays a fundamental crisis in computational biology [3]. In molecular biology, the high-throughput technologies have generated a huge sum of data about the function, structure, and evolution of biological macromolecules at the genome level. To identify the interactions of these molecules in the cell, it is essential to classify the molecules into significant categories that are associated with the existing biological data [4]. It is essential to classify the

N. Sajithra (✉) · D. Ramyachitra · P. Manikandan
Department of Computer Science, Bharathiar University, Coimbatore, India
e-mail: sajithramidhun@gmail.com

D. Ramyachitra
e-mail: jaichitra1@yahoo.co.in

P. Manikandan
e-mail: manimkn89@gmail.com

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_10

protein structures for comparing the protein structures and also it can be used to easily identify and group similar folds and families of protein structures. One of the major advantages of classifying the protein structures is to introducing some wisdom of order to the increasing volume of available protein structural data. The obstacle occurs in classifying the protein structures is the reality that the structures are frequently made up of distinct globular proteins [5].

In recent days, the computational techniques in structural genomics are well found that proteins often have similar folds, even if their protein sequences are apparently dissimilar [6]. A prototype computational platform named Homologous Structure Finder (HSF) has the capability of incorporating a variety of comparison methods and parameterizations. HSF platform has been validated by more complex classification tasks and it successfully performed an automatic comparison and classification of the set of 3D protein structures [7]. Most of the common methods are used for comparing and classifying the virus proteins based on sequence information [8]. Nowadays, it is easier to identify the similar proteins that share a similar function based on their structure pretty than on their sequences [9]. In this article, the new emerging techniques and tools to classify the protein structures are studied. The increasing protein structure database helps both biological and computational scientists to predict structural similarity, functions, and progress the diagnosis and treatment of diseases. The remaining sections of the paper are organized as follows. Section 2 deals with the challenges in the structural classification of proteins. The algorithms which are used to classify the protein structures are discussed in Sect. 3. Section 4 provides the tools involves in classifying the protein structures. Section 5 discusses the protein structure databases. The performance measures are discussed in Sect. 6 and conclusion is provided in Sect. 7.

2 Challenges in Protein Structure Classification

The major challenge in structural biology is to classify well-known protein structures to replicate their evolutionary, structural, and functional relatedness. The relationships between the protein structures are personified as hierarchical classification schemes in Structural Classification of Proteins (SCOP) [10]. But, there are some challenges in classifying the protein structures in a hierarchical manner [11]. On the other hand, the lower levels have a propensity to be dense and overlapping to contain the ubiquitous functional similarity and evolutionary relationships. In order to organize the complex universe of known protein structures, human expertise has been playing a crucial role. For a major number of proteins, there are differences in classification among any two release versions of SCOP and a huge number of proteins remain listed under the Not a True Class level. Hence, an enthusiastic significance is needed in the automation of protein structure classification [12].

3 Techniques for Protein Structure Classification

There are several "-omics" projects such as genomics, functional genomics [13], and structural genomics [14] that are producing a huge amount of information and it has been stored into databases. One of the key successes in human ability is to analyze and organize the biological information. In other words, the human ability is to classify the proteins based on the sequence, structure, and functions to perform classification. Some of the techniques which are used to perform the protein structure classification are shown in Fig. 1.

The automated classification of protein structure is performed by using different algorithms from different categories of supervised algorithms on a set of 11,336 pairs of protein domains with 35% sequence identity [15].

To analyze the [16] protein structures Gauss metrics method is used and this protein will be formed as clusters. The advancement of Gauss metric is scaled Gauss metric that is used to maintain the database and it will provide the automatic classification. This method has high speed and it provides better performance.

In the Protein Data Bank [17], the graph-based classification method provides the primary overview of the entire protein complex and allows nonredundant sets to be derived at diverse levels in detail. This reveals that among one-half and two-thirds of recognized protein structures are multimeric, depending on the stage of accepted redundancy. The protein structures in provisions of the topological arrangement of their subunits are analyzed and find that they form a tiny number of arrangements compared with all theoretically possible ones.

The 3D shape-based approach [18] is tested with three classification methods to examine the performance of this approach.

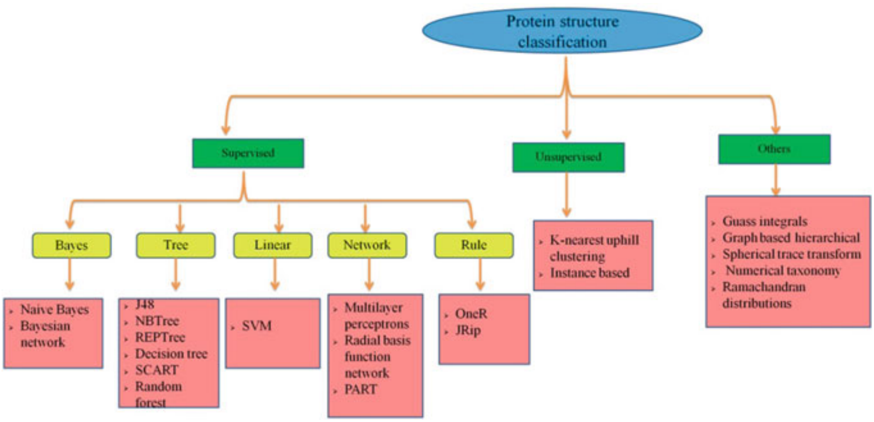


Fig. 1 Classifications of protein structure techniques

A novel density-based K-nearest uphill clustering technique can efficiently reduce noisy pairwise structure similarities of proteins and identifies the density peaks as cluster centers [19].

The illustration of the crisis [20] in an attribute-based method allows the purpose of numerous well-established machine learning algorithms.

The usability of the random forest algorithm [21] in classifying larger domains is verified by applying it to domains consisting of four, five and six Secondary Structure Elements.

A numerical taxonomy is introduced for automatic classification of protein structures and identification of evolutionary relationships [22]. The topology of protein space is explored using structural similarity. Penetrating for clusters of structural neighbors of proteins, where the members constantly share numerous functional attributes directs to an optimal partitioning of protein structure space. This clustering matches up well to the analog or homolog boundaries drained by biologists, with applications in the creation of functional hypotheses in nonhypothesis-driven structural genomics efforts.

A nonparametric technique is developed for the combined estimation of numerous bivariate density functions for a group of populations of protein backbone angles [23].

4 Tools

Protein structure classification algorithms and tools are mainly used for identifying the structural similarities between the proteins. In recent years, the algorithms which are used for automatically identifying the structural relationships have been a sizzling area in computational biology [24]. Numerous techniques are used for handling the shifts in protein secondary structure orientations and the wide indels among distant homologs such as COMPARER [25], DALI [26] and SSAP [27]. Several resources are available for classifying the protein and its families such as HAMAP [28], TIGRAMs [29], PANTHER [30], SFLD [31], PFAM [32], SMART [33], InterPro [34], PRINTS [35], CATH [36], NCBI-CDD [37], ECOD [38], and SCOP [39].

5 Databases

Several databases has been used for analyzing the problem of protein structural classification that includes Protein Data Bank (PDB) [40], Universal Protein Resource [41], Database of Secondary Structure of Protein (DSSP) [42], Structural Classification of Proteins (SCOP) [10], SCOP2 (a successor of SCOP) [43] and Class, Architecture, Topology, Homology (CATH) [36]. From the above databases, SCOP and CATH have become precious resources in protein structural classification research.

6 Performance Metrics

In general, the performance measures are chosen based on the algorithms used in that research work. Some of the common measures that can be used in protein structure classification algorithms are Accuracy, True Positive rate (TP), False Positive rate (FP), Mathews Correlation Coefficient (MCC), F-Measure and it can be calculated by using Eqs. 2, 3, 4, 5 and 6, respectively. The overall accuracy is the percentage of correctly predicted proteins and it is calculated by using Eq. 1.

$$\text{Overall Accuracy} = \frac{\text{Number of correctly predicted proteins}}{\text{Total number of proteins in the database}} \tag{1}$$

$$\text{Accuracy} = \frac{(\text{TP} + \text{TN})}{(\text{TP} + \text{FP} + \text{FN} + \text{TN})} \tag{2}$$

$$\text{True Positive/Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{3}$$

$$\text{False Positive} = \frac{\text{FP}}{\text{FP} + \text{TN}} \tag{4}$$

$$\text{Matthews Correlation Coefficient} = \frac{\text{TP} \times \text{TN} - \text{FP} \times \text{FN}}{\sqrt{(\text{TP} + \text{FP})(\text{TP} + \text{FN})(\text{TN} + \text{FP})(\text{TN} + \text{FN})}} \tag{5}$$

$$\text{F-Measure} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \tag{6}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{7}$$

where,

True Positive (TP)	included class objects with corrected
False Positive (FP)	incorrectly included class objects with uncorrected
True Negative (TN)	excluded class objects with corrected
False Negative (FN)	excluded class objects with uncorrected

ROC analysis is used to compare the rate of True Positive as a function of the rate of False Positive and it is scored with the area under the corresponding curve. The ROC value of 1 indicates that all TPs are identified first and this match to the ideal measure [9].

7 Conclusion

Performing an automated classification of protein structures remains a challenging problem in computational biology and it is necessary to develop an accurate method for classifying the protein structures from the structural data. Some of the techniques which are used to perform the protein structure classification are discussed in this review article. And also this article provides a brief note on tools and databases, which can be used to perform protein structure classification. Finally, this article discussed some common performance metrics for evaluating the performance of protein structure classification algorithms. Protein structure classification is still an attractive research area with a bunch of interesting possibilities for additional improvements in terms of efficiency and accuracy.

Acknowledgements The authors like to thank the Department of Science and Technology (DST), New Delhi (DST/INSPIRE Fellowship/2015/IF150093) for the financial support under INSPIRE Fellowship for this research work.

References

1. Richardson J (1981) The anatomy and taxonomy of protein structure. Adv Protein Chem 34:167
2. Branden C, Tooze J (1991) Introduction to protein structures. Garland Publishing, New York
3. Kolodny R et al (2013) On the universe of protein folds. Annu Rev Biophys 42:559–582
4. Ouzounis CA et al (2003) Classification schemes for protein structure and function. Nat Rev Genet 4(7):508–519
5. Hadley C, Jones DT (1999) A systematic comparison of protein structure classifications: SCOP, CATH and FSSP. Structure 7(9):1099–1112
6. Pastore A, Lesk AM (1990) Comparison of the structures of globins and phycocyanins: evidence for evolutionary relationship. Proteins 8(2):133–155
7. Ravantti J et al (2013) Automatic comparison and classification of protein structures. J Struct Biol 183(1):47–56
8. Palmenberg et al (2009) Sequencing and analyses of all known human rhinovirus genomes reveal structure and evolution. Science 324:55–59
9. Le Q et al (2009) Structural alphabets for protein structure classification: a comparison study. J Mol Biol 387(2):431–450
10. Murzin AG et al (1995) Scop: a structural classification of proteins database for the investigation of sequences and structures. J Mol Biol 247:536–540
11. Govindarajan S et al (1999) Estimating the total number of protein folds. Proteins: Struct Funct Bioinform 35:408–414
12. Andreeva et al (2008) Data growth and its impact on the SCOP database: new developments. Nucleic Acids Res 36:D419–D425
13. Burley S et al (1999) Structural genomics: beyond the human genome project. Nat Genet 23:151–157
14. Hieter P, Boguski M (1997) Functional genomics: it's all how you read it. Science 278:601–602
15. Jain P et al (2009) Supervised machine learning algorithms for protein structure classification. Comput Biol Chem 33(3):216–223
16. Røgen P, Fain B (2003) Automatic classification of protein structure by using Gauss integrals. Proc Natl Acad Sci U S A. 100(1):119–124
17. Levy ED et al (2006) 3D complex: a structural classification of protein complexes. PLoS Comput Biol 2(11):e155
18. Daras P et al (2006) Three-dimensional shape-structure comparison method for protein classification. IEEE/ACM Trans Comput Biol Bioinform 3(3):193–207
19. Cui X, Gao X (2017) K-nearest uphill clustering in the protein structure space. Neurocomputing 220:52–59
20. Leon F et al (2009) Performance analysis of algorithms for protein structure classification. In: 2009 IEEE 20th international workshop on database and expert systems application. https://doi.org/10.1109/dexa.2009.17. ISBN: 978-0-7695-3763-4

21. Jain P, Hirst JD (2010) Automatic structure classification of small proteins using random forest. BMC Bioinform 11:364
22. Dietmann S, Holm L (2001) Identification of homology in protein structure classification. Nat Struct Biol 8(11):953–957
23. Najibi SM et al (2017) Protein structure classification and loop modeling using multiple Ramachandran distributions. Comput Struct Biotechnol J 8(15):243–254
24. Swindells MB et al (1998) Contemporary approaches to protein structure classification. BioEssays 20(11):884–891
25. Sali A, Blundell TL (1990) Definition of general topological equivalence in protein structures. A procedure involving comparison of properties and relationships through simulated annealing and dynamic programming. J Mol Biol 212:403–428. https://doi.org/10.1016/0022-2836(90)90134-8
26. Holm L, Sander C (1993) Protein structure comparison by alignment of distance matrices. J Mol Biol 233:123–138. https://doi.org/10.1006/jmbi.1993.1489
27. Taylor WR, Orengo CA (1989) Protein structure alignment. J Mol Biol 208:1–22
28. Pedruzzi I et al (2013) HAMAP in 2013, new developments in the protein family classification and annotation system. Nucleic Acids Res 41:D584–D589
29. Haft DH, Selengut JD, White O (2003) The TIGRFAMs database of protein families. Nucleic Acids Res 31:371–373
30. Mi H, Muruganujan A, Thomas PD (2013) PANTHER in 2013: modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees. Nucleic Acids Res 41:D377–D386
31. Akiva E et al (2013) The structure–function linkage database. Nucleic Acids Res 42:D521–D530
32. Finn RD et al (2014) Pfam: the protein families database. Nucleic Acids Res 42:D222–D230
33. Letunic I, Doerks T, Bork P (2015) SMART: recent updates, new developments and status in 2015. Nucleic Acids Res 43:D257–D260
34. Hunter S et al (2012) InterPro in 2011: new developments in the family and domain prediction database. Nucleic Acids Res 40:D306–D312
35. Attwood TK et al (2012) The PRINTS database: a fine-grained protein sequence annotation and analysis resource—its status in 2012. Database 2012:bas019
36. Sillitoe I et al (2015) CATH: comprehensive structural and functional annotations for genome sequences. Nucleic Acids Res 43:D376–D381
37. Marchler-Bauer A et al (2013) CDD: conserved domains and protein three-dimensional structure. Nucleic Acids Res 41:D348–D352
38. Cheng H et al (2014) ECOD: an evolutionary classification of protein domains. PLoS Comput Biol 10:e1003926
39. Andreeva A et al (2007) Data growth and its impact on the SCOP database: new developments. Nucleic Acids Res 36:D419–D425
40. Bernstein FC et al (1977) The protein data bank. Eur J Biochem 80:319–324
41. Consortium, U (2008) The universal protein resource (UniProt). Nucleic Acids Res 36:D190–D195
42. Kabsch W, Sander C (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers 22:2577–2637
43. Andreeva A et al (2014) SCOP2 prototype: a new approach to protein structure mining. Nucleic Acids Res 42:310–314

Emotion Analysis Through EEG and Peripheral Physiological Signals Using KNN Classifier

Shourya Shukla and Rahul Kumar Chaurasiya

Abstract Emotions are the characteristics of human beings which are triggered by the mood, temperament or motivation of an individual. Emotions are nothing but the response to the stimuli that are experienced by the brain. Any changes in one's emotional state results in changes in electrical signals generated by the brain. The emotions can be explicit or implicit, i.e. either emotion may be expressed or remain unexpressed by the individual. As these emotions are experienced by the individual as the result of the brain stimulus, we can observe Electroencephalogram (EEG) signal to classify the emotions. Some of the physiological signals may also be taken into account as any change in emotional state result in some physiological changes. For the analysis, we have used the standard DEAP dataset for emotion analysis. In the dataset, the 32 test subjects are shown with 40 different 1-minute music videos and the EEG and other physiological signals are recorded. On the basis of the Self-Assessment Manikins (SAM), we classify the emotion state in the valence arousal plane. The K-Nearest Neighbour classifier is used to classify the multi-class emotions as higher/lower levels of the valence arousal plane. The comparison of KNN with other classifiers depicts that KNN has produced best average accuracy of 87.1%.

1 Introduction

Electroencephalogram is the electrical signals produced by the brain. EEG signal are used to analyse the emotional state of the subject under study. Emotions are triggered in the brain by stimuli such as video signals or audio signals etc., [1]. Depending upon the mood, temperament and motivation of the subject, different emotions are experienced. The brain works in different frequency bands, namely delta, theta, alpha, beta and gamma [2, 3]. Beta (12–30 Hz) band is emitted when someone is conscious

S. Shukla (✉) · R. K. Chaurasiya (✉)
National Institute of Technology, Raipur, India
e-mail: shourya.shukla2000@gmail.com

R. K. Chaurasiya
e-mail: rkchaurasiya@nitrr.ac.in

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_11

and alert, and the subject is thinking or concentrating. Alpha band (8–12 Hz) is most active when the subject is in the state of physical and mental relaxation but in aware and conscious. Theta (4–7 Hz) band is associated with daydreaming or sleepy state. It is also called the creative state. The delta band (0.1–4 Hz) is the lowest frequency state related to deep sleep. At the upper highest frequencies, that is the gamma band (>30 Hz) the subject id in deep meditation, mainly found significant in Buddhist monks.

The human–human interaction is easier as human are more aware of the sentiments of the human they are interacting, but in the case of the Brain–Computer interaction (BCI), it is more complicated as the computer is unaware of the emotions of the human subject. Here, we present a method which uses a signal-based approach by extracting information from the EEG and peripheral physiological signals. These peripheral physiological signals are useful in predicting the emotional state of the subject as they complement EEG signals in the emotional analysis. These peripheral signals are significant in different frequency bands and hence, the Discrete Wavelet Transform (DWT) of the signals is taken into consideration. The DWT of a signal splits the signal into the higher frequency detail (D) and the lower frequency approximation (A) coefficients [2].

The features are extracted from the wave decomposed signals, and the KNN classifier is used to classify the emotions as low valence low arousal, low valence high arousal, high valence low arousal and high valence high arousal. The valence indicates the pleasing level of the brain. A high valence level indicates happy or elated emotion, whereas a low valence indicates sad or stressed behaviour. The arousal indicates activeness of the brain. A higher arousal indicates alert or excited response, whereas the lower level of arousal indicates uninterested or bored response of an individual.

2 Related Works

The EEG signals were first recorded by the English scientist Richard Caton [4] in the year 1875. The study of the EEG signals was first explored by Hans Berger in 1920 [4]. The first study of emotion and physiological signals goes back to the year 1941 by Hadley, J. M. in which the author described the relationship between the EEG and peripheral physiological signals while performing multiplications of varying difficulties [5]. Plutchik, R first associated with high sound intensities on the performance feelings and physiology of the subject [6]. Moon et al. proposed the method for video preferences based on the extracted using EEG-based responses quadratic–discriminant–analysis-based model using BP features [7]. There are many proposed ways to classify the emotion of the subject using EEG. Ekman and Friesen, 1987 were the first to propose the six emotions existing that can be classified using facial signals [8]. The valence arousal plane is used to separate different emotional state as proposed by Sander Koelstra et al. (2011). A Database for Emotion Analysis using Physiological Signals (DEAP dataset) uses the EEG and other physiological signals to classify the emotions in the valence arousal plane using F1 score and naïve Bayes classifier is performed [9].

The peripheral physiological signals play an important role in determining the emotion as they complement the EEGs and provide the information about the subject's reaction to a stimuli. Torres-Valencia et al. suggested a multimodal emotion recognition using the DEAP dataset by Hidden Markov Model (HMM) using the Galvanic Skin Resistance (GSR) and Heart Rate (HR) [10]. Ramasamy et al. defined the heart–brain interaction through the EEG and ECG involving the emotion through biofeedback system [11]. Li et al. related the EEG with peripheral physiological signals such as Electrooculogram (EOG), Electrocardiogram (ECG), Electromyogram (EMG), skin temperature variation and electrodermal activity in a brain–computer interface system to measure the attention level in ubiquitous environment [12].

3 Materials and Methods

3.1 Dataset for Experimental Analysis

In the last few years, the BCI is one of the most studied topics in the field of machine learning. The DEAP dataset is being used for research purpose by many scholars. We are also using the DEAP dataset which is available for the research work. An End-User License Agreement (EULA) is acquired to access the data. The dataset consists of the records of 32 patients (subjects). Out of 32 subjects, 22 subjects' facial recording is provided. Each subject's file consists of data and label file. The data file has the 32-channel EEG recordings along with 8 peripheral physiological signals. The physiological signals which are used to understand the emotional state of the subject are Galvanic Skin Resistance (GSR), Respiration Amplitude (RA), Skin Temperature (ST), Electrocardiogram (ECG), Blood Volume Pressure (BVP) and Electromyogram (EMG) of the zygomaticus and trapezoidal muscles. The eyeball movement is captured by the Electrooculogram (EOG) signal. The signals are sampled at the rate of 512 Hz and the peripheral physiological are further downsampled to 256 Hz. To record the EEG and other signals, visual stimuli are used. 40 videos were selected using the web-based survey and the stimulating 1 minute of each video is shown to the subject. The data is hence of the dimension $40 * 40 * 8064$, for 40 min, 40 signals for each minute and 8064 samples of each signal. 22 subjects' facial video was also recorded using SONY DCR-HC27E camcorder.

The EEG signals were obtained using a 10–20 electrode system. The EEG has 32-channel electrode having a 10–20 system with odd number of electrode placed on the left hemisphere and electrode with the number placed on the right hemisphere of the brain. The dataset used the Bio Semi Active Two System for recording the EEG signals.

The labelling of the data is done by the rating given by the individual. The process is named Self-Assessment Manikins (SAM). In the labelling process, after completion of every 1-min video, the subjects are asked to rate the video between 1 and 9, where 1 being the lowest and 9 being the highest rating. The subject is asked to move the cursor horizontally and click on the rating bar to give the scores on the arousal and valence parameters. To define a binary class system, midpoint threshold were taken on the rating scale of 1–9 for arousal and valence. Other SAMs like dominance, liking and familiarities were also assessed. The dominance rating represents the feeling of being empowered, whereas for liking scale, thumbs up or thumb down symbols were used. Familiarity rating suggests that how well the subject knows or remembers the video stimulus. The labelling of data is done as high valence high arousal, high valence low arousal, low valence high arousal and low valence low arousal. Thus, multi-class classification is performed on basis of these four classes.

3.1.1 Preprocessing, Feature Extraction and Classifier Methodology

The DEAP dataset contains records of 32 test subjects, and the large dataset makes it difficult to work on the integrated data. We segmented the data for each minute recording, i.e. time signals were obtained for each minute video. Therefore, each subject's psychophysiological signals are divided into 40 segments. Then for each minute, we find the discrete wavelet transform coefficient. These coefficients are time signals with ascending order of frequencies. The length of each frequency band is given in the length coefficient matrix. The DWT gives the details and approximations as response to high-pass and low-pass filter, respectively. The five-level DWT is obtained for the EEG signals of the frequency range 0–30 Hz. The five-level DWT is also used for the peripheral physiological signals as the EMG signals works on the higher frequency range of 4–40 Hz.

3.1.2 Feature Extraction Using DWT

EEG being a non-stationary signal, Fourier transform is not a suitable transform for it. Hence, to find the feature matrix, five-level DWT is used. The DWT decomposes the wavelets in the time–frequency domain. The DWT decomposes the signal according to the increasing frequency bands. The frequency decomposition is obtained by selecting odd and even samples of the signal and then, these samples are passed through low-pass and high-pass filter, respectively, further, the filtered signals are downsampled by a factor of 2. The resulting high-pass downsampled signals is the detail signals and low-pass downsampled signals are the approximate signals. The multilevel DWT is performed by applying recursively DWT to the $(n-1)$th approximate samples [13]. For each minute video, the five-evel DWT is applied on the psychophysiological signals, hence each signal is decomposed into five detail levels D1–D5 and an approximate A5, features are then extracted and classified by KNN classifier. As different signals works in different frequency ranges (EEG in the 0–30 Hz, whereas trapezoidal muscles in 4–40 Hz), all the DWT coefficients are taken into account.

Feature extraction is performed on the decomposed wavelets to study the EEG and other physiological signals in different frequency ranges. The EEG features extracted for classification of the emotional state of the subject are mentioned below:

(1) First feature consists of the logarithmic values of power spectral densities of EEG samples of mean values of each band.
(2) Second feature consists of the difference in power spectral densities of corresponding right and left hemispheric electrode of the EEG signal.

The GSR measures the resistance of the skin. It is related to the amount of perspiration the subject has while watch the video stimulus. The degree of perspiration provides information about the nervousness and anxiety of the subject. The resistance decreases with increase in perspiration. Lang et al. suggested that the arousal is correlated with the mean value of GSR [14]. The following features are extracted from the GSR signal:

(1) The average values of each band D1–D5 and A5 are extracted.
(2) The mean values of derivative of GSR is extracted for each band.
(3) The mean values of derivative for negative values only. Hence, average decrease rate during perspiration is evaluated.
(4) Ratio of number of negative sample to the total number of samples in each band.
(5) Number of local minima of each band.
(6) Zero-crossing rate of the GSR.

The BVP is measured using plythesmograph. It is the measure of the pressure by which heart exerts pressure into the arteries. The plythesmogarph is attached to thumb of the subject. The BVP can be used to find the heart rate and heart rate variability. The heart rate and its variability are correlated to the emotional state as the faster heart rate indicates subject under stress. The higher blood pressure indicates the sense of fear or surprise. The following are the feature extracted from the blood volume pressure of the subject undergoing visual stimuli:

(1) Average value of BVP.
(2) Standard deviation of BVP.
(3) Average value and standard deviation of the heart rate (heart rate is identified with the help of local maxima, i.e. the heart beat and the heart beats per minute is the heart rate).
(4) Heart rate variability (Iwona Cygankiewicz et al., suggested that HRV reflects beat to beat changes in R-R interval. Heart rate changes may occur as a response to the physical and mental stress [15]. And hence, related to the emotional state of the subject under study).
(5) Inter-beat interval.
(6) Energy of the blood volume level in each frequency band.
(7) Energy ratio of blood volume level in consecutive frequency bands.
(8) Power spectral densities of blood volume level.

The next physiological signal is the RA. The RA is recorded as the speed of respiration depends on the emotional state, a slow RA represents relaxation and fast or irregular respiration implies a feeling of fear or anger. The RA has high correlation with the arousal. The feature extracted from RA is as follows:

(1) Inter-band difference in energy of respiration signal.
(2) Average respiration signal.
(3) Mean of derivatives of each frequency band.
(4) Standard deviation of RA. (This feature shows variation in respiration amplitude which in turn shows the change in mood of the patient).
(5) Range or greatest breath time taken by the subject.
(6) Spectral centroid of the respiration frequency bands.
(7) Breathing rate.
(8) Power spectral density of respiration amplitude.

ST is also used to extract the emotion and hence the following features are extracted:

(1) Average value of each band of temperature.
(2) Average of its derivatives are extracted to find the mean rate of change of ST
(3) Power spectral densities

Muscle signals were obtained in the DEAP dataset to extract information about the facial expression and shoulder movement. The facial muscle plays an important role in expressing one's emotions. Hence, EMG of zygomaticus muscle was recorded. The shoulder muscle movement is implies laughter (happy) emotion or anger as implicit tags. The features extracted are:

(1) Energy of the muscle signals.
(2) Mean of the EMG.
(3) Variance of the zygomaticus and trapezoidal muscles.

The eyeball movement and tracking is involved in predicting one's emotions. The blinking rate is decreased to a large extent when the person experiences high arousal. The EOG signal can be related to anxiety of the person also. The features presented by the DEAP dataset are:

(1) Energy of the EOG signals in each band.
(2) Mean of the EOG signal in each frequency band.
(3) Variance in eyeball movement.
(4) Blinking rate of the eye. (The blinking rate is determined by the detectable peaks of the EOG).

3.2 Classification Using KNN Classifier

The KNN is a supervised learning algorithm, hence, it works on the given dataset directly. For any new instance, predictions are made on the search of the K most proximate instances. Here in the case of KNN, these instances are known as "Neighbours". To measure the parameter of the neighbour, we are using the Euclidean distance. The K-nearest neighbour's Euclidean distance is calculated with respect to the instance data X. The majority voting is performed and the instance X is allotted the class of majority of K-nearest neighbour's class [16, 17].

4 Experiments and Results

In this paper, we have used a supervised machine learning approach to classify EEG and peripheral physiological signals to differentiate between different emotional states. The EEG and peripheral physiological signals were wavelet decomposed to five-level DWT using Haar wavelets. Thirty-four features are extracted using the detail and approximate subbands of the decomposed wavelets. Vectors are formed by concatenating these features. This feature matrix is classified using KNN classifier with fivefold cross-validation method. This method is proposed to find the accuracy of the emotional analysis using psychophysiological signals.

Different pilot experiments were done and $K = 10$ was chosen based on the performance in terms of accuracy.

The accuracy is computed using the confusion matrix. The multi-class confusion matrix is $4 * 4$ matrix with true positive are placed at $i = j$ cell of A_{ij} confusion matrix. The sensitivity and specificity is calculated using confusion matrix.

$$\text{SENSITIVITY} = \frac{\text{TRUE POSITIVE}}{\text{TRUE POSITIVE} + \sum_N \text{FALSE NEGATIVE}} \tag{1}$$

$$\text{SPECIFICITY} = \frac{\text{TRUE NEGATIVE}}{\text{TRUE NEGATIVE} + \sum_N \text{FALSE POSITIVE}} \tag{2}$$

The multi-class accuracy is obtained from the sensitivity and specificity of the classified results:

$$\text{ACCURACY} = \frac{\text{SENSITIVITY} + \text{SPECIFICITY}}{2} \tag{3}$$

The four classes are classified with the KNN classifier with 87.1% accuracy.

In Fig. 1, the confusion matrix of KNN classifier with data points is inserted as true class versus predicted class is shown. The diagonal elements of the $4 * 4$ matrix depicts that the data points are accurately predicted, whereas the other elements of the matrix depicts the missclassification of the data points.

In Fig. 2, the confusion matrix is shown in terms of percentage of data points accurately classified or misclassified. In Fig. 3, the region of convergence is shown for the KNN classifier.

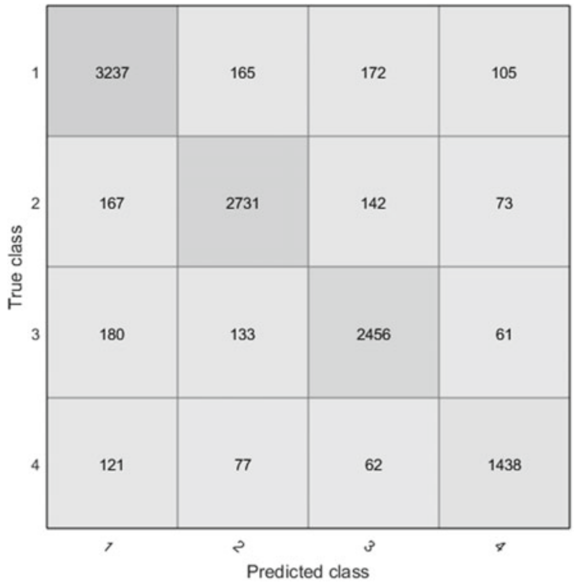


Fig. 1 Confusion matrix in terms of data points of KNN classifier

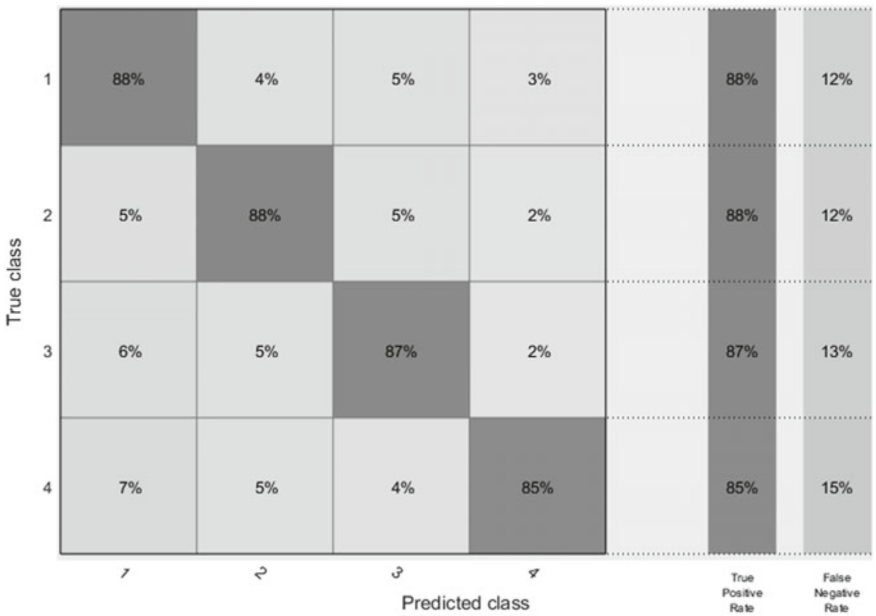


Fig. 2 Confusion matrix in terms of percentage accuracy of KNN classifier

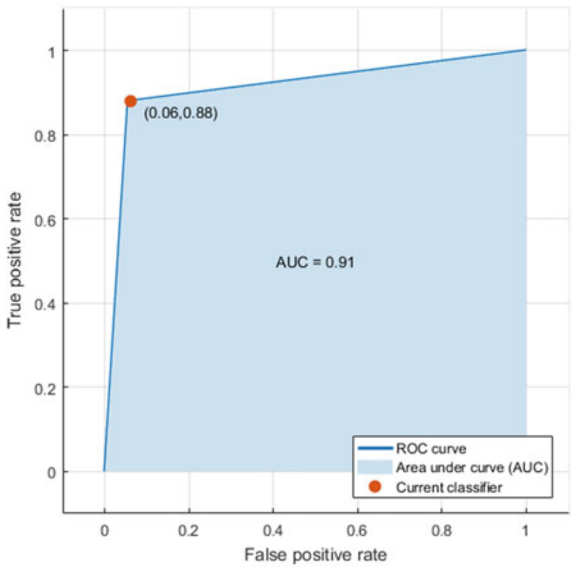


Fig. 3 Region of convergence curve for KNN classifier

5 Conclusion

In this paper, we have proposed a DWT-based approach to the emotion analysis of the humans using video stimuli. The proposed approach trains the psychophysiological signals using the five-level DWT and then, extracting features such as mean, standard deviation, power spectral density, average decrease rate during decay time, heart rate and heart rate variability and eye blinking rate to classify the emotion expressed explicitly or even implicit emotions are detected. The KNN classifies the emotions as into four classes that are high valence high arousal, high valence low arousal, low valence high arousal and low valence low arousal. A good classification accuracy of 87.1% is obtained using KNN classifier suggest that this algorithm is suitable for training data to classify the emotions using visual stimuli.

Use of weighted KNN with feature reduction using principal component analysis will also be tried to reduce the system complexity. In future, the focus will be to use deep learning, Random forest etc., to further improve the performance.

References

1. Dasdemir Y, Yildirim E, Yildirim S (2017) Analysis of functional brain connections for positive–negative emotions using phase locking value. Cogn Neurodyn 11(6):487–500
2. Jahankhani P, Kodogiannis V, Revett K (2006) EEG signal classification using wavelet feature extraction and neural networks. In: International symposium on modern computing, 2006. JVA'06. IEEE John Vincent Atanasoff 2006. IEEE, New York, pp 120–124

3. Başar E, Başar-Eroglu C, Karakaş S, Schürmann M (2001) Gamma, alpha, delta, and theta oscillations govern cognitive processes. Int J Psychophysiol 39(2):241–248
4. Haas LF (2003) Hans Berger (1873–1941), Richard Caton (1842–1926), and electroencephalography. J Neurol Neurosurg Psychiatry 74(1):9
5. Hadley JM (1941) Some relationships between electrical signs of central and peripheral activity: II. During 'mental work'. J Exp Psychol 28(1):53
6. Plutchik R (1959) The effects of high intensity intermittent sound on performance, feeling and physiology. Psychol Bull 56(2):133
7. Moon J, Kim Y, Lee H, Bae C, Yoon WC (2013) Extraction of user preference for video stimuli using EEG-based user responses. ETRI J 35(6):1105–1114
8. Ekman P, Friesen WV, O'sullivan M, Chan A, Diacoyanni-Tarlatzis I, Heider K, Scherer K (1987) Universals and cultural differences in the judgments of facial expressions of emotion. J Pers Soc Psychol 53(4):712
9. Koelstra S, Muhl C, Soleymani M, Lee JS, Yazdani A, Ebrahimi T, Pun T, Nijholt A, Patras I (2012) Deap: a database for emotion analysis; using physiological signals. IEEE Trans Affect Comput 3(1):18–31
10. Torres-Valencia CA, García HF, Holguín GA, Álvarez MA, Orozco Á (2015) Dynamic hand gesture recognition using generalized time warping and deep belief networks. In: International symposium on visual computing. Springer, Cham, pp 682–691
11. Ramasamy M, Varadan VK (2017) Study of heart-brain interactions through EEG, ECG, and emotions. In: Nanosensors, biosensors, info-tech sensors and 3D systems 2017, vol 10167. International Society for Optics and Photonics, p 101670I
12. Li Y, Li X, Ratcliffe M, Liu L, Qi Y, Liu Q (2011) A real-time EEG-based BCI system for attention recognition in ubiquitous environment. In: Proceedings of 2011 international workshop on ubiquitous affective awareness and intelligent interaction. ACM, New York, pp 33–40
13. Hoa LT, Anh ND (2007) Orthogonal-based wavelet analysis of wind turbulence and correlation between turbulence and forces. Vietnam J Mech 29(2):73–82
14. Lang PJ, Greenwald MK, Bradley MM, Hamm AO (1993) Looking at pictures: affective, facial, visceral, and behavioral reactions. Psychophysiology 30(3):261–273
15. Cygankiewicz I, Wranicz JK, Bolinska H, Zaslonka J, Zareba W (2004) Relationship between heart rate turbulence and heart rate, heart rate variability, and number of ventricular premature beats in coronary patients. J Cardiovasc Electrophysiol 15(7):731–737
16. Zhang ML, Zhou ZH (2007) ML-KNN: a lazy learning approach to multi-label learning. Pattern Recogn 40(7):2038–2048
17. Parthasarathy G, Chatterji BN (1990) A class of new KNN methods for low sample problems. IEEE Trans Syst Man Cybern 20(3):715–718

Enhancement of Optical Coherence Tomography Images: An Iterative Approach Using Various Filters

M. Saya Nandini Devi and S. Santhi

Abstract Speckle is an important noise in the optical coherence tomography (OCT) images that play a vital role in the degradation of the visual quality of images and makes it difficult to assess the quality of the images. In order to improve the quality, filtering is essential which removes the noises and reproduce the OCT images. This proposed work addresses various filters like Mean, Median, Adaptive Median, Gaussian, and Wiener in order to enhance the OCT image. This paper gives estimate of various filtering techniques on the basis of the performance indices calculated from the experimental results. The results of this research work suggest the filter suitable for OCT images depending on the Cross-Correlation, Mean Square Error-(MSE) and Peak Signal-to-Noise Ratio-(PSNR) which are very much used as performance indices in medical image applications and its analysis. Initially, the filter is tested with standard medical image and performance measures are calculated. Next, noise-corrupted OCT image is applied as an input to the filter for the purpose of denoising and finally, a comparative analysis was performed.

1 Introduction

Speckle is a typical noise in engineering that corrupts the details of the image. It is the initial factor that restricts contrast resolution of images, thereby restraining detectability of small, low-contrast objects and makes the images usually complex for interpreting by nonspecialist. This noise affects the visual impact of the optical coherence images and leads to a problem in diagnosing the images. Speckle noise is most prevalent among tomography images and is acquired by the pixels during the scanning or imaging phase of the image acquisition Phenomena. The incident and

M. Saya Nandini Devi · S. Santhi (✉)
Department of Electronics and Instrumentation Engineering, Annamalai University, Chidambaram, India
e-mail: santhi.sathyamurthy@gmail.com

M. Saya Nandini Devi
e-mail: nandini46@gmail.com

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_12

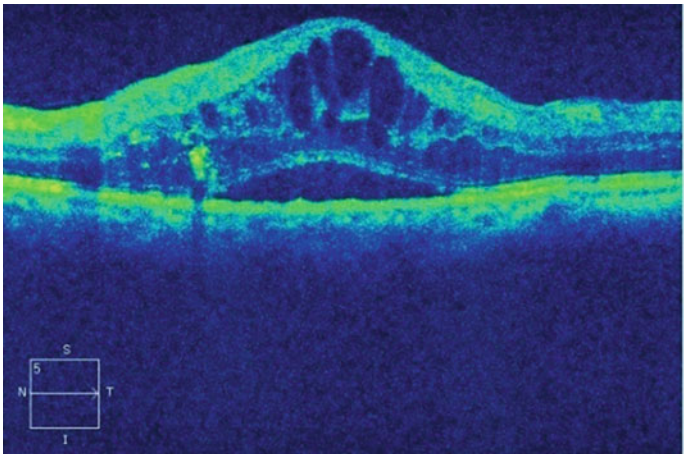


Fig. 1 Sample OCT image

reflected beams of the imaging devices cause a backscattering effect on the pixels of the image and appear as tiny speckles of noise particles which severely degrade the quality of the image and even causing loss of critical data if not detected and treated early. It is a high speed, high-resolution three-dimensional clinical diagnostic images in guiding the various procedures in the medical field. The coherence imaging modalities suffer from the speckle noise.

Speckle noise is similar to salt and pepper noise except that it is multiplicative in nature, unlike its counterpart which is additive in nature. This multiplicative noise causes increased grainy appearance due to its multiplicative nature and should be filtered early during the preprocessing phases. By assuming the linear polarization and speckle pattern which is sophisticated, speckle perhaps modeled as multiplicative noise that follows unit mean Rayleigh distribution. Figure 1 depicts the randomly selected medical image and the sample of OCT image. Analysis algorithms, 2-D display, 3-D volume rendering, and image processing effective applications are restricted by this. This requires reduction without affecting features of OCT images. To carry out the parallel valuation of speckle reduction algorithms on the basis of quality estimate is the key objective of this paper.

2 Related Work

A survey has been made to find the filter performance in OCT images. Literature [1] depicts the function of the 3D median filter in the filtering process of the noise in an OCT image. The author addresses the median filter as the widely used filter in image processing as it has the power to eliminate the noise in the edges and reduces the blurring effects particularly for the images of large dimensions. In this article, the median filter is applied for reducing the speckle noise and further, it was processed for the histogram intensity equalization. Literature [2] describes the use of noise-adaptive wavelet thresholding filter in the noise removal process in OCT images. The work has concluded with the performance evaluation of SNR and preserving structural by the quantity β, compared to conventional wavelet domain thresholding and Gaussian filtering. The finding from the article [3] describes the various denoising filters for decreasing speckle noise, in the dermatology B scans with edge preserving potential. The results are obtained by the author using the Block matching and enhanced sigma filters which are considered as a 2D filter in the denoising. The research work [4] depicts the attributes of the speckle reduction in the area of interpretation and diagnosis of spectral optical coherence retinal image. In the work, anisotropic diffusion filtering has been executed by measuring the parameters of structural similarity index for preserving the edges in that image. Literature [5] describes the de-speckling of Multiframe Optical Coherence Tomography based on anisotropic diffusion for similarity comparison between frames.

Article [6] depicts the work of spatial technique for speckle reduction in OCT. On the basis of finding the most likely intensity values at certain locations, the algorithm has been presented. The algorithm presented is based on finding the most likely intensity value for each pixel at a specific location in the image. The research paper [7] presented method basing on the scattering model for the calculation of attenuation coefficient in OCT profiles. The phantom experiments used in this have the ability for estimating accurately attenuation coefficients in case of uniforms and even layered. The literature [8] depicts the function of Wavelet denoising filter in the filtering process of multiframe OCT data. The wavelet coefficients detail is scaled with the weights, averaged, and transformed back. Literature [9] describes the Non-Local Mean [NLM] filter with double Gaussian anisotropic kernels removes speckle noise within OCT image. In this research, filters are compared with the various filters namely the NLM, Median, Bilateral, and the Wiener filters. Literature [10] investigated speckle noise as well as motion artifact and described their relation. Article [11] describes Gray Level CO-occurrence Matrix (GLCM) feature extraction and Back-Propagation Neural Network (BPNN) classification is used for reduction of speckle noise in an OCT image. The result shows better performance in the noise identification. Literature [12] detailed the Multi-Photon Tomography (MPT) and OCT systems to acquire the deviant skin lesions and the 3D images for the normal. The purpose of the paper is visualizing the 3D morphology of various layers of the skin and to detect characteristic features with the help of MPT and OCT. The research paper [13] presents an automatic segmentation of retinal layers in OCT images in case of spectral domain with the help of sparsity-based denoising, graph theory, S-GTDP, and the support vector machines. Literature [14] describes repetitious advancement on the basis of bilateral filtering in order to decrease the speckle in case of multiframe OCT data. Conventional bilateral filter applied for enhancing the OCT data and after that, the bias caused by the noise is decreased from every frames that are filtered. In [15], Multiscale Sparsity Based Tomographic Denoising (MSBTD) algorithms remove the speckle noise in an SDOCT image. This technique illustrates efficiently the simultaneous reduction of noise and interpolation of missing data.

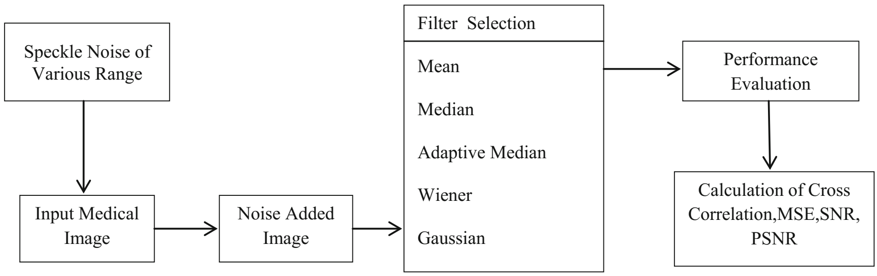


Fig. 2 Proposed model flow diagram

3 Proposed Work

The proposed work investigates the performance of five different filters employed for decreasing speckle noise of OCT image. Speckle noise and its interference were calculated based on the general model and it is given as

$$Dn(nf) = c(nc)\cdot\ p(np) + s(ns) \tag{1}$$

where Dn is the observed data with noise, p is multiplicative, and s is additive component of speckle noise in n data that affects noiseless constant data set. To handle data in the speckle noises, the components are generally decoupled into logarithmic space. The noisy image in the process contains the pixel values which are much greater than the constant input image which already used as a reference. To estimate the noise parameters, the patches and the pixels are calculated so that a correlation coefficient value is obtained. The filers which are used in the proposed model reduce the noises based on the estimation process. In general, the Gaussian noise is close to multiplicative component of speckle. Hence, the distribution of the speckle can be related as

$$D(nf) = N(nf|\gamma) = \frac{a}{\sqrt{n\pi\alpha_i}} \mathrm{e}^{-n(f-1)\frac{n}{2\alpha^2}} \tag{2}$$

where nf is deviation factor in the noise present in the entire image (Fig. 2).

3.1 Wiener Filter

Wiener estimates the local mean and the variance of a sliding window of size $n \times m$ pixels and each pixel is located on the ith row and the kth column of the image and it generates a new pixel consisting of estimated pixel values and it is calculated as

$$f(i,k) = \mu + \frac{\sigma^2 - \beta^2}{\sigma^2}(f(i,k) - \mu) \tag{3}$$

where μ is the mean, σ^2 is the variance and β^2 is the local variance of the sliding window of $n \times m$ pixels and each pixel is located in the ith row and the kth column of the image and it generates the pixel value as

$$\beta^2 = \frac{1}{mn}\left\{\frac{f^2 \times M}{mn} - \frac{f \times M}{mn}\right\}^2 \tag{4}$$

where M indicates the matrix of ones with the same dimension as the sliding window indicating the convolution operation.

3.2 Median Filter

This is the most used nonlinear filter for removing the speckle noise and other noises and it preserves the other details in the image. The filter works by replacing the pixel of an image by median value belonging to the pixel with nearest neighbors. Mathematical model of this filter is given as

$$f_{i,k} = \text{med}\big(n\big(G_{i,k}\big)\big) \tag{5}$$

where $n\big(G_{i,k}\big)$ gives the neighborhood function with M centered at pixel location i, k.

3.3 Adaptive Median Filter

The Adaptive Median Filter is an advanced method of filtering process compared to standard median filter. The filter performs the spatial process to determine affected pixels in the image by comparing the pixel values with its neighbor pixels. The noise pixel was identified by observing the majority of pixels and other different pixels. These noises are replaced by median pixel values of neighborhood values. The process of adaptive median filter is described as follows

Level A: A1 = Zmed – Zmin

A2 = Zmed – Zmax

if A1 > 0 *AND A2* < 0, *go to level B*

else increase the window size

fwindowsize $<$ *Smax, repeatlevelA*

else output Zxy

Level B: B1 = Zxy – Zmin

B2 = Zxy – Zmax

if B1 > 0 *AND B2* < 0, *output Zxy*

else output Zmed

3.4 Gaussian Filter

The Gaussian filter function is used in the research area as for defining the probability distribution of the noise and the possible values give the space which varying the values from negative to positive. A two-dimensional Gaussian function is generally used for the filtering process and the filter works based on point spread function. Using convolution, the 2D Gaussian function performs the distribution function. Distribution verge close to zero around three standard deviations away from the mean. 99% belonging to the distribution comes close to the three standard deviations. It implies that the general restriction of the kernel size is containing values not beyond three standard deviations close to the mean. Kernel coefficient sampled from 2D Gaussian function is given as

$$G(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2 - y^2}{2\sigma^2}} \tag{6}$$

An effective yet simple method of filtering out noise components from an image is the averaging or means filtering which is simple in construction and consumes least computational time and complexity. It usually utilizes a kernel or mask-based approach of an image $m \times n$ which is moved over the entire image in a horizontal or vertical scanning manner. The filter works on a replacement basis where the noisy pixel is replaced by the mean of the surrounding pixels resulting in the filtered image. However, the PSNR could be visualized as to be of a very high value as the replaced pixel represents only the averaged values of the nearby pixels.

3.5 Mean Filter

The Average or Mean Filter is an easy sliding window spatial filter which substitutes centered value inside the window by "average" of every pixel values inside window.

$$\text{Meanfilter}(x_1 \dots x_n) = \frac{1}{N}\sum_{t=1}^{N} x_i \tag{7}$$

3.6 Performance Evaluation

The common metrics are the SNR and the PSNR of the filtered image. The PSNR of an image is given by

$$\text{PSNR} = 10 \times \log_{10} \frac{255^2}{\text{MSE}} \tag{8}$$

where "MSE" is the mean square error and is given by

$$\text{MSE} = (\text{sum}(\text{sum}(\text{err.} * \text{err})))/(M * N) \tag{9}$$

The implementation of the proposed method is as follows:

- Initialize the filter and generate variable correlation coefficient values based on speckle noises added in the original medical image.
- Correlation value for speckle of n values = corr2(Original Image, Filtered Image)
- $\text{MSE} = (\text{sum}(\text{sum}(\text{err.} * \ \text{err})))/(M * N)$
- $\text{SNR} = 10 * \log 10((1/M * N) * \text{sum}(\text{sum}(A * B))/(\text{MSE}))$
- $\text{PSNR} = 10 * \log 10(255 * 255/\text{MSE})$
- Compare the values of MSE, SNR, PSNR
- Repeat the previous steps for OCT image and compare the performance of filters.

4 Results and Discussion

The proposed speckle noise reduction process is tested on a 1.7 GHz processor with 2 GB RAM running Windows 7 and coded by Matlab 15. An OCT image is taken into account for the identification of suitable filter for the removal of speckle noise. The process starts from developing Matlab code for the various filters such as Mean filter, Median, Adaptive Median, Wiener, and Gaussian filter and analyzing their performance for a test image. The test image considered is MRI axial neck image and speckle noise with amplitude ranging from 10 to 100% was added to it. The noise added images were passed as input to the filters to identify a better filter for the removal of speckle noise. The performance measures such as correlation coefficient, mean square error, SNR, and PSNR were calculated for the test image to validate the filter algorithms. It has been observed that the Gaussian filter performs greatly for MRI image. Next, the filter algorithms were tested with OCT image that naturally has speckle noise embedded in it the process of its acquisition. The performance measures were again calculated to identify a suitable filter for the removal of speckle noise. It is observed that for an OCT image wiener filter performs well in removing speckle noise and is proposed for our future work of developing detection and classification of macular diseases on using OCT images (Figs. 3, 4, 5 and 6; Tables 1 and 2).

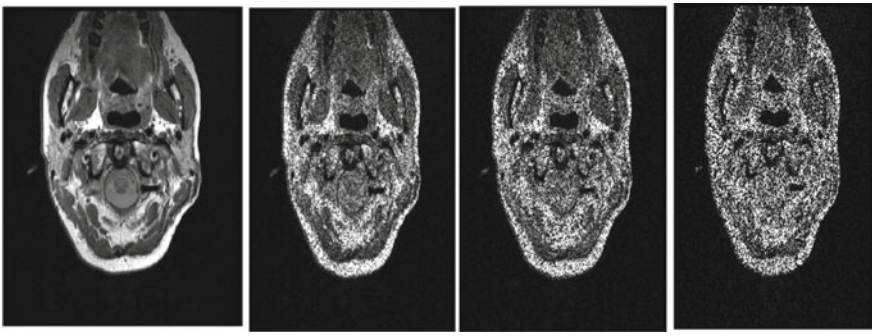


Fig. 3 Results obtained with special noise added for various value to medical image (Axial neck MRI): (1) Original image, (2) 10% speckle noise added, (3) 60% speckle noise added, (4) 100% speckle noise added

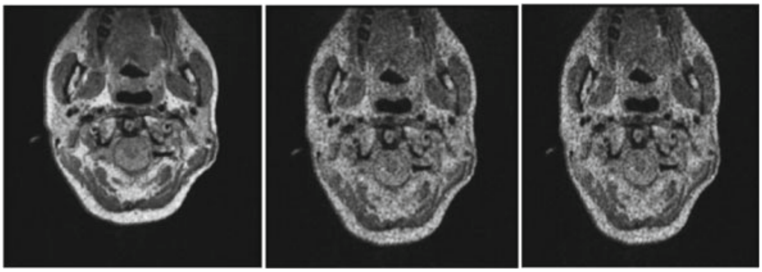


Fig. 4 Results acquired with Gaussian filter when applied for the medical image (Axial neck MRI): (1) 10% filtered output, (2) 60% filtered output, (3) 100% filtered output

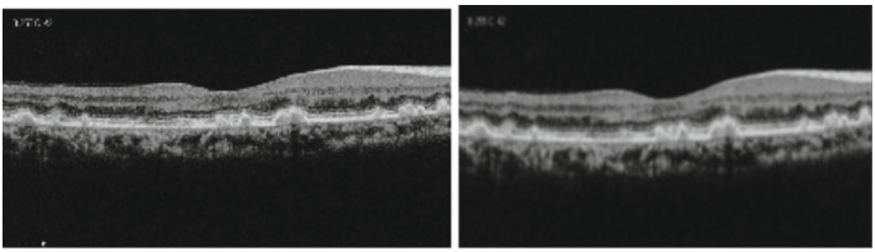


Fig. 5 Results acquired with Gaussian filter when applied for the OCT image: (1) Speckle noise image; (2) Filtered output image

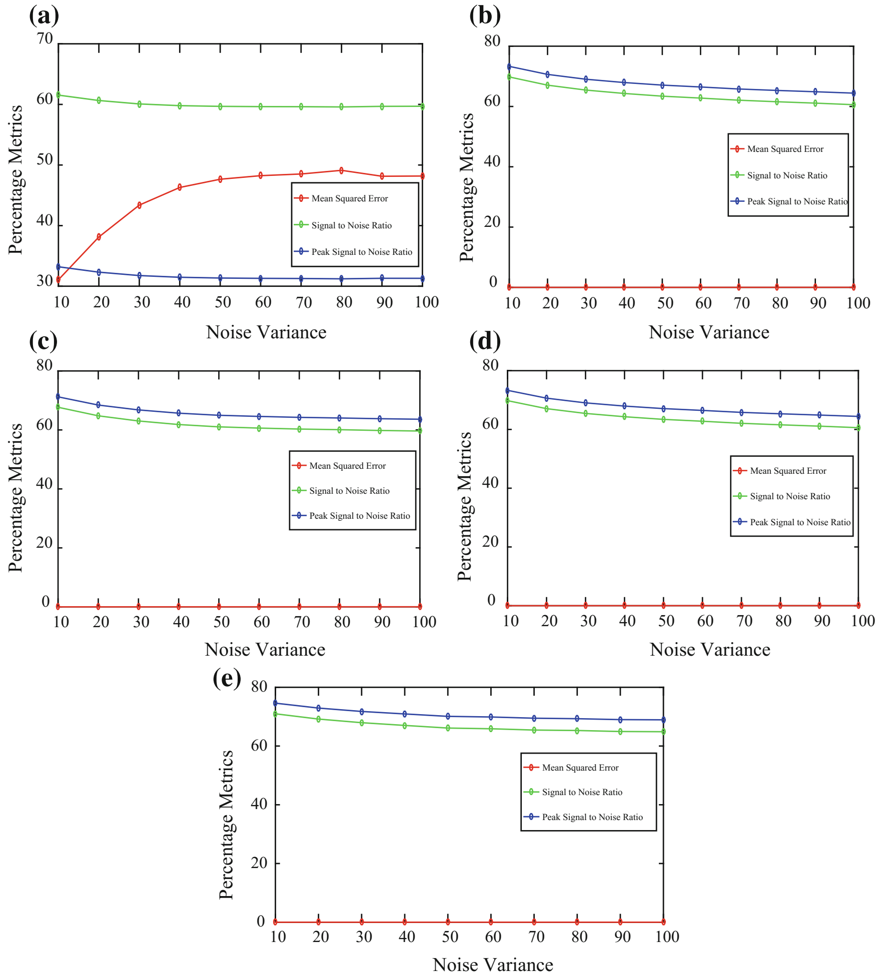


Fig. 6 Performance metrics for filters and its comparison for medical image (Axial Neck MRI): **a** Mean filter **b** Median filter **c** Adaptive median filter **d** Wiener filter **e** Gaussian filter

Table 1 Comparison of cross-correlation values for the input medical image

Filters	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
Mean	0.97	0.95	0.93	0.92	0.90	0.89	0.87	0.86	0.84	0.83
Median	0.94	0.90	0.86	0.83	0.82	0.80	0.80	0.79	0.79	0.78
Adaptive median	0.98	0.97	0.96	0.95	0.94	0.93	0.93	0.92	0.91	0.91
Wiener	0.96	0.92	0.89	0.86	0.84	0.82	0.81	0.80	0.79	0.78
Gaussian	0.98	0.97	0.97	0.96	0.95	0.95	0.95	0.94	0.94	0.93

Table 2 Performance measurement values for the input OCT image

Filters	MSE	SNR	PSNR
Mean	0.0018	70.4935	75.5302
Median	0.0024	69.3073	74.3760
Adaptive median	5.6705e−04	75.6791	80.5946
Wiener	4.0974e−04	77.0844	82.0058
Gaussian	0.0025	69.0584	74.1606

5 Conclusion

This research paper examined various filter performance characteristics for processing OCT images in order to remove speckle noise that is inherently present in these type of images. Algorithms for various filters such as mean, median, adaptive median, Wiener, and Gaussian were developed through Matlab coding. The performance of the various filters was first validated for the normal medical image and then the filter algorithms were applied to optical coherence tomography image. The process started with the speckle noise addition at various levels to the normal medical image and then the performance measures were calculated. Then, the various filters were applied with OCT images which contained speckle noises and performance measures such as SNR, PSNR, and Mean square error were calculated. It is observed that Gaussian serves better for medical image and Wiener filter is superior for OCT image.

Acknowledgements We would like to thank Dr. M. Ravishankar, director of the Nethralayam and Senior Consultant of the Rajan Eye Care Hospital, Chennai for providing the Optical Coherence Tomography Images.

References

1. Gocławski J, Sekulska-Nalewajko J (2016) A new idea of fast three-dimensional median filtering for despeckling of optical coherence tomography images. Image Process Commun 20(3):25–34
2. Zaki F, Wang Y, Su H, Yuan X, Liu X (2017) Noise adaptive wavelet thresholding for speckle noise removal in optical coherence tomography. Biomed Opt Express 8(5):2720
3. Gómez-Valverde JJ, Ortuño JE, Guerra P, Hermann B (2015) Evaluation of speckle reduction with denoising filtering in optical coherence tomography for dermatology. IEEE, vol 2, No 1, pp 494–497
4. Padmasini N, Abbiraime KS, Yacin SM (2014) Speckle noise reduction in spectral domain optical coherence tomography retinal images using anisotropic diffusion filtering. IEEE, vol 1, No 3, pp 244–248
5. Bian L, Suo J, Chen F, Dai Q (2015) Multiframe denoising of high-speed optical coherence tomography data using interframe and intraframe priors. J Biomed 20(3)
6. Avanaki MR, Cernat R, Tadrous PJ, Tatla T, Podoleanu AG, Hojjatoleslami S (2013) Spatial compounding algorithm for speckle reduction of dynamic focus OCT images. IEEE Photonics Technol Lett 25(15):1439–1442

7. Vermeer K, Mo J, Weda J, Lemij H, de Boer J (2014) Depth-resolved model-based reconstruction of attenuation coefficients in optical coherence tomography. Biomed Opt Express 5(1):322–337
8. Mayer MA, Borsdorf A, Wagner M, Hornegger J, Mardin CY, Tornow RP (2012) Wavelet denoising of multiframe optical coherence tomography data. Biomed Opt Express 3(3):572–589
9. Aum J, Kim J, Jeong J (2015) Effective speckle noise suppression in optical coherence tomography images using nonlocal means denoising filter with double Gaussian anisotropic kernels. Appl Opt 54(13):D43–D50
10. Rajabi H, Zirak A (2016) Speckle noise reduction and motion artifact correction based on modified statistical parameters estimation in OCT images. Biomed Phys Eng Express 2
11. Mittal A (2013) Automatic noise identification using GLCM properties. Int J Adv Res Comput Sci Softw Eng 3:943–947
12. Alex A, Weingast J et al (2013) Three dimensional multiphoton/optical coherence tomography for diagnostic applications in dermatology. J Biophotonics 6(4):352–362
13. Srinivasan PP et al (2014) Automatic segmentation of up to ten layer boundaries in SD-OCT images of the mouse retina with and without missing layers due to pathology. J Biomed Opt Express 5(2):348–365
14. Sudeep PV, Niwas SI, Palanisamy P, Rajan J, Xiaojun Y, Wang X, Luo Y, Liu L (2016) Enhancement and bias removal of optical coherence tomography images: an iterative approach with adaptive bilateral filtering. Comput Biol Med 71:97–107
15. Fang L, Li S, McNabb RP, Nie Q, Kuo AN, Toth CA, Izatt JA, Farsiu S (2013) Fast acquisition and reconstruction of optical coherence tomography images via sparse representation. IEEE Trans Med Imaging 32(11):2034

Hybrid Method for Copy-Move Forgery Detection in Digital Images

I. J. Sreelakshmy and Binsu C. Kovoor

Abstract Digital image authenticity is significant in many social areas. Image forgery detection becomes a challenging task. Copy-move forgery is one of the tampering techniques which is frequently used, part of the image is copied and pasted to other parts of the same image. This paper proposes a new method for copy-move forgery detection. Proposed method integrates both block-based and keypoint-based forgery detection. Host image is first divided into blocks and keypoints are extracted from each image block. Blocks are compared based on the keypoints in them. Number of similar keypoints identified from a pair of blocks exceeds a preset threshold, then those block pair is matched. Matched blocks are considered as the forged region and Output is displayed after neighbour pixel merging and morphology operations. The accuracy of the method is calculated and analysed with different images.

1 Introduction

In this digital world, digital images play a crucial role. Majority of the information is shared in digital image form. Every second, millions of images are uploaded and shared through social media. Digital images are now considered as a valid evidence even in the courts. Though we are very much depending on the images, the credibility of these images is still a question. Because of the widespread availability of photo editing tools, a person without any technical skill can modify the images. Manual analysis is infeasible on this huge volume of images. So forensic tools are needed to be more advanced to check the integrity of the digital images.

Forgery detection mechanisms are mainly of two types, active and passive. Active methods like Watermarking, Digital signature need to embed information at the time of creation, then only it can be verified in the future. Passive method, on the other

I. J. Sreelakshmy · B. C. Kovoor (✉)
Cochin University of Science and Technology, Kochi, India
e-mail: binsu.kovoor@gmail.com

I. J. Sreelakshmy
e-mail: 93sreelakshmy@gmail.com

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_13

hand, can verify digital images without any prior knowledge about the image. So it does not need advanced equipment, as well as the quality of the image, does not get degraded because of the data embedding as in Active method. These advantages make the recent researches tend to be more on to Passive forgery detection area.

One of the most common passive forgery detections is copy-move image forgery detection. In Copy-move forgery, some part of the image is copied and pasted to other parts of the same image. Copy-move forgery can easily hide any objects in the scene or it can create multiple clones of a particular object in the image, thus the meaning of the image gets altered entirely after forgery. Existing copy-move forgery detection techniques can be categorized into two, keypoint-based and block-based methods. Here, we are discussing a hybrid method of the above two. Host image is first divided into blocks as in block-based method and like keypoint-based method, keypoints are used for block matching. Matched blocks are considered as the forged region and exact regions are extracted after neighbour pixel analysis and morphology operations. The accuracy of the method is calculated and analysed with different variations of the image also.

The remaining contents of the paper are presented in the following manner. Next section deals with the previous work related to copy-move image forgery detection. Section 3 explains the proposed method, Sect. 4 contains Result and Analysis, and in the end, we have conclusion and references.

2 Related Work

Existing Copy-move forgery detections can be either block-based or keypoint-based method. In block-based methods, the host image is divided into segments and for each segment, block features are computed. Pair of blocks whose feature vector similarity exceeds the threshold value is identified as forged regions. In keypoint-based methods, keypoints are extracted from the image and feature descriptor is computed for every keypoint. These feature descriptors are compared to find duplicated regions.

Initially, most of the works are done using block-based methods. All of them describe the same workflow, differ only in the method chosen to compute the block feature vector. Fridrich et al. [1] proposed extraction of Discrete Cosine Transform(DCT) coefficients from each block for feature vector and checking the segment pairs with same DCT coefficients detected as suspected regions. Principal Component Analysis (PCA) is used by Popescu and Farid [2], in this paper, blocks with similar PCA components identified as forgery regions. Mahdian and Saic [3] uses Blur moment invariants with PCA reduction. After the overlapped segmentation of the image, for each block, a feature vector is constructed with 24 blur moment invariants. Principal component transformation is used to reduce the feature vector dimension. The similarity of the blocks is then identified by comparing the feature vectors. K-d tree is used for an efficient searching neighbour blocks. If the neighbourhood of two similar blocks is also similar then those pairs are labelled.

Luo et al. [4] proposed intensity components as a feature vector. The image is divided into small overlapping blocks and characteristic features (7 intensity components) of each block is extracted. Comparing the feature vector of each block with one another and blocks which have the same shift vector and the difference in their characteristic coefficients is less than the threshold value are marked as similar blocks. Another variation of DCT method is discussed in [5]. DCT applied to each block and feature reduction by circular block representation and matching reduced 4 vector feature set for forgery detection. DWT can also use to compute the block feature vector [6]. DWT is applied to the image and select only the low-frequency portions from DWT result. Low-frequency regions are segmented and apply DCT on each segment. DCT coefficient features extracted from each block and lexicographically sorted and find the correlation coefficient to match the features. Bayram et al. proposed Fourier Transform coefficients as block feature vector [7]. But all of these block-based methods are not robust to any type of transformations and compression. Also, the complexity gets large as the image size increases.

In order to overcome the drawbacks of block-based methods keypoint-based methods are introduced. Because of the transformation invariant property of the keypoints, it can detect the duplicated regions even after scaling and rotation. Huang et al. proposed Scalar Invariant Feature Transform (SIFT) algorithm for keypoint extraction [8]. SIFT keypoints are extracted from the image. Forgery is detected by matching the keypoints. Match value exceeds the given threshold then those points are marked as forged ones. Bo et al. [9] followed the same procedure but instead of SIFT, they use Speeded-Up Robust Features (SURF) algorithm to retrieve the keypoints. An improved keypoint method is introduced by Ardizzone et al. [10], keypoint extraction as well as segmentation are done for the detection. After the keypoint extraction, image is segmented to triangles based on these keypoints. Inner angles, colour and mean vertex descriptors are compared to find similar triangles and thereby detecting forgery. Finally, RANSAC filter is applied to remove false matches. Keypoint-based methods cannot detect the forgery if the region is too small or forged regions are homogeneous. We cannot find enough keypoints in those cases [11].

3 Proposed Method

To detect copy-move forgery in digital images, we introduce an improved hybrid procedure of block-based and keypoint-based method. So we can make use of the advantages of both the methods. First, the image is segmented into overlapping blocks as in block-based method, then keypoints are extracted using SURF algorithm from each block. And we are comparing these keypoint descriptors to detect forgery region as in keypoint-based method. Finally, apply some morphology operations and adding similar neighbour to get a more accurate region. The basic workflow is depicted in Fig. 1.

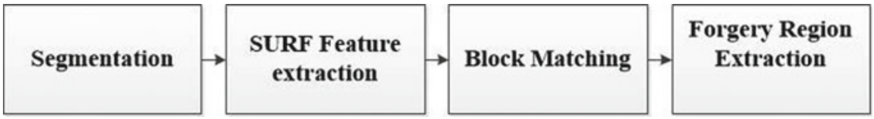


Fig. 1 Work flow of the proposed model

3.1 Segmentation

Simple Linear Iterative Clustering (SLIC) algorithm is used to segment the image into overlapping blocks. Even though SLIC is a well-known k-mean clustering-based algorithm for image segmentation, its segmentation is so much depending on the superpixel size value inputted by the user. So the user must be careful while selecting the superpixel size value. To get a perfect segmentation, for different images its suitable superpixel size value also differs. Our method computes the superpixel value adaptively, thereby reducing the extra burden on the user. First examine the texture of the image, if it is a smooth image, pixel size will be set as a large value, or set a small value if the image is a detailed one. Existing relation between texture and frequency of the image help us to understand whether the host image is smooth or detailed one. An image with the majority of the frequency energy account of low-frequency energy considered as a smooth image, otherwise detailed image. DWT using Haar wavelets is used to calculate the percentage of low-frequency energy in an image and thereby computing the corresponding superpixel size. With this superpixel size value, SLIC segmentation can be carried out [12].

Haar wavelet transform is applied in 4 levels. Low-frequency energy is calculated from the approximate coefficient results from the fourth stage of DWT. Diagonal, vertical, and horizontal coefficients of all levels are used to find out the high-frequency energy.

$$E_{\mathrm{LF}} = \sum |\mathrm{CA}_4| \tag{1}$$

$$E_{\mathrm{HF}} = \sum_i \sum |\mathrm{CD}_i| + \sum |\mathrm{CH}_i| + \sum |\mathrm{CV}_i|, \tag{2}$$

where CA_4—Approximation coefficients at the fourth level of DWT;

CD_i, CH_i and CV_i—Detailed coefficients at the ith level of DWT, $i = 1, 2, \ldots, 4$.

Percentage of low-frequency energy (P_{LF}) can be computed easily from this E_{LF} and E_{HF} values and based on the obtained P_{LF} value we can set superpixel size accordingly, using the following equations:

$$P_{\mathrm{LF}} = \frac{E_{\mathrm{LF}}}{E_{\mathrm{LF}} + E_{\mathrm{HF}}} \cdot 100\% \tag{3}$$

$$S = \begin{Bmatrix} \sqrt{0.02 \times M \times N} & P_{\mathrm{LF}} > 50\% \\ \sqrt{0.01 \times M \times N} & P_{\mathrm{LF}} \leq 50\% \end{Bmatrix} \tag{4}$$

3.2 SURF Keypoint Extraction

After segmenting the images using the computed superpixel value, the second phase is the extraction of keypoints from each of these image segment. SIFT and SURF are the commonly used algorithms for keypoint extraction. The accuracy of SLIC is slightly greater than SURF. But SURF can perform as good as SIFT in small regions as well as it is faster than SIFT. Here we have to extract the keypoints not from the entire image but from the individual segments only. Because of the small image regions, it is better to use SURF. So SURF algorithm is used here for keypoint extraction. SURF is carried out mainly in a two steps, Keypoints detection and calculate descriptor for each identified keypoints. These descriptors are used in the next stage of the matching process.

3.3 Block Matching

To detect duplicated regions, we have to find out the similar blocks. Similarity measure of all possible pair of blocks is computed and the block pair whose similarity measure exceeds the preset threshold $\mathrm{T_B}$ is considered as suspected regions. Similarity measure of a block pair is the number of matching keypoints in them. Keypoints are matched based on its descriptors. If the number of keypoints matched among two blocks greater than T_{B}, and the distance between them is greater than threshold T_{D} then we have marked those keypoints as suspected points. Neighbour segments of a uniform portion of the image naturally have some similar keypoints and because of this, they are wrongly detected as a suspected region. Distance threshold T_{D} will remove this type of false positives.

3.4 Forgery Region Extraction

At the end of block matching, we got all the duplicated points. We have to extract the region of forgery from these marked points. The image is segmented with the same SLIC algorithm with a larger superpixel size, in order to make very small segments. Now the segments containing marked feature points are considered as suspected regions of forgery. There may be chances that, to miss out some of the forged pixels because of the segmentation procedure. So check the neighbour pixels of these suspected regions. Merge the neighbour pixels with similar characteristics. Similar pixels are identified here by checking the average of RGB values of each pixel and for any pair of pixels, if the difference in average is not more than preset threshold T_{NP}, we treat them as similar characteristic pixel pair. So the neighbour pixels are compared in this way and the similar characteristics neighbours are also merged into the suspected region. Resulting forgery regions are displayed after some morphologic operations.

4 Result and Analysis

A series of experiments were conducted to evaluate the effectiveness and robustness of the proposed image forgery detection scheme. In the following experiments, the image dataset given in [13] is used and the proposed method was tested, implemented in MATLAB R2013a. Figure 2 shows the test image and output obtained using the proposed method.

The accuracy of the detection result is evaluated at the pixel level based on its ground truth image. Main characteristics used for accuracy analysis are precision and recall. The probability that the detected regions are relevant is termed as Precision and probability that the relevant regions are detected is Recall. According to the randomness of test images, F-score is used to test accuracy. F-score is calculated using the following equation:

$$F1\text{score} = 2 \cdot \frac{\text{precison} \cdot \text{recall}}{\text{precision} + \text{recall}} \tag{5}$$

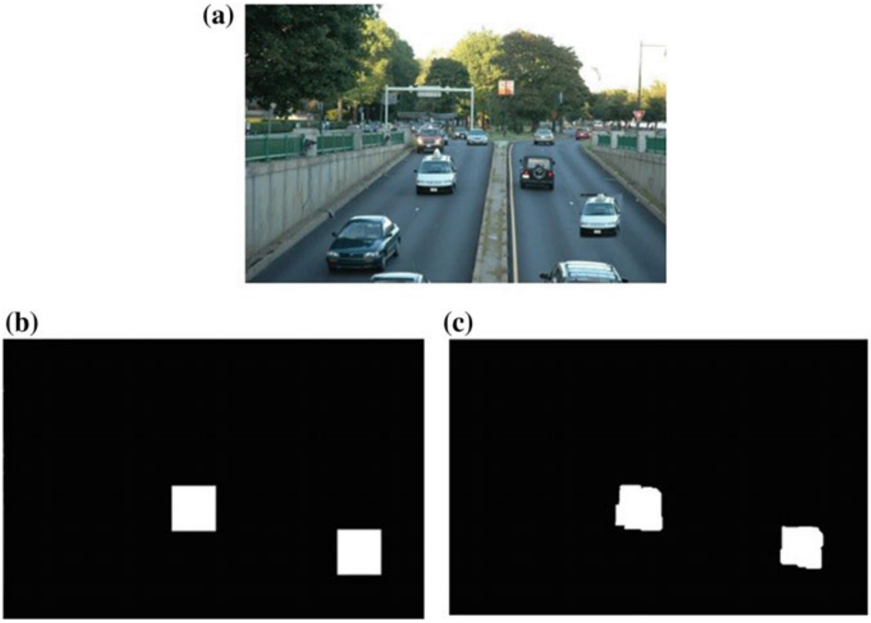


Fig. 2 **a** Input image **b** Ground truth of corresponding image **c** Forgery detection output

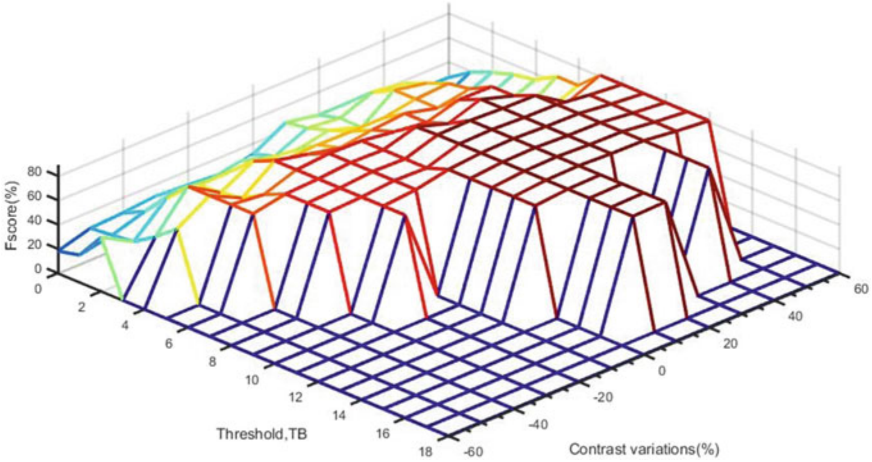


Fig. 3 F-score value against contrast variations and block matching threshold

4.1 *Effect of Contrast Variations*

A total of 30 images were tested by varying the contrast. The F-score of the test corresponding to the variation of contrast between −60 and +60% and block matching threshold (T_B) values between 0 and 18 is given in Fig. 3. The F-score is found to be minimum for all values of contrast where the threshold is less than or equal to 0 and greater than or equal to 18. F-score is positive over contrast between 0 and 40% when the threshold is 10–16. But in low-contrast images, we get the F-score value only in small values of threshold (between 2 and 6). This indicates that threshold and contrast significantly influence the performance of the model. On high-contrast versions of the image, we have got more accurate results, if we set threshold to be a large value. Similarly, threshold should be set minimal to process low-contrast images. In real time, if the model is used for outdoor and the detection is to be made for only certain hours of the day, the threshold can be approximately set corresponding to the contrast level of the scenery to be detected.

The proposed model uses keypoints to identify the duplicated regions. Analysis of regions will be better if there are more number of keypoints. The number of keypoints obtained in each contrast level is observed in Fig. 4. As the contrast increases, the number of keypoints extracted by the SURF algorithm also increases. So high-contrast images are more suitable for this method whereas low-contrast image results with less accuracy because they did not get enough keypoints for the evaluation. To make low-contrast images to work well in this method, we have to do a preprocessing stage, which increases the percentage of contrast level. Then, we can detect the forgeries more accurately in low-contrast images also.

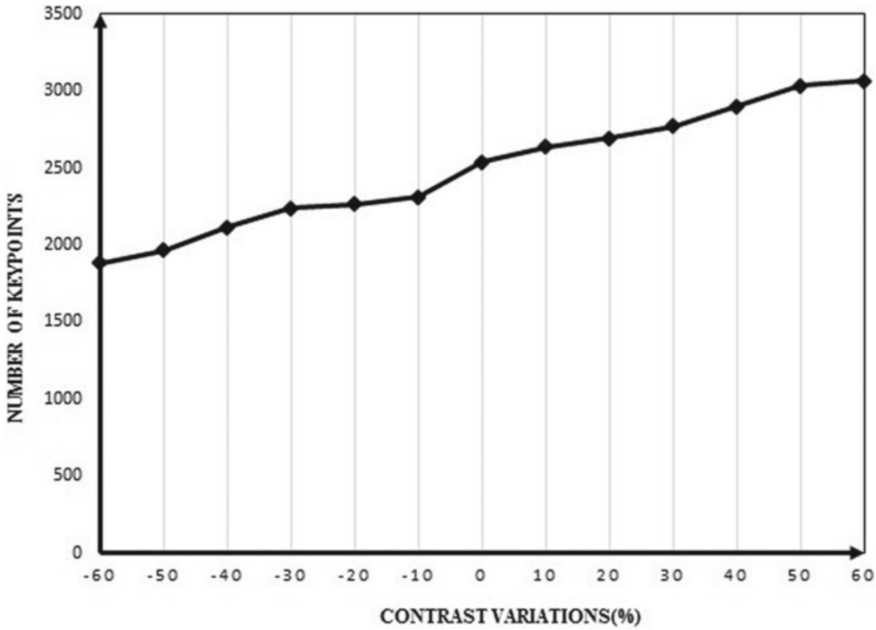


Fig. 4 Relation between contrast variation and number of keypoints

5 Conclusion

In this digital era, digital images are easily tampered and the detection of forgeries are becoming more difficult. An improved method to detect copy-move forgery is discussed in this paper. Host image is divided into overlapping blocks and keypoints are extracted from each block. To detect the duplicated regions within the image, we have to find similar blocks. All pair of blocks is compared based on the keypoints in them, and the matched blocks are identified as suspected regions. To get a more accurate result similar neighbour pixel merging and morphologic operations are also performed on the suspected regions. Model is analysed for different types of images and discovered various ranges of the threshold to detect the forgery in real time.

References

1. Fridrich AJ, Soukal BD, Lukas AJ (2003) Detection of copy-move forgery in digital images. In: Proceedings of digital forensic research workshop
2. Popescu AC, Farid H (2004) Exposing digital forgeries by detecting duplicated image regions. Department Computer Science, Dartmouth College, Technology Report TR2004-515
3. Mahdian B, Saic S (2007) Detection of near-duplicated image regions. Comput Recogn Syst 2:187–195

4. Luo W, Huang J, Qiu G (2006) Robust detection of region-duplication forgery in digital image. In: 18th international conference on pattern recognition, vol 4. IEEE, New York
5. Cao Y (2012) A robust detection algorithm for copy-move forgery in digital images. Forensic Sci Int 214(1):33–43
6. Hayat K, Qazi T (2017) Forgery detection in digital images via discrete wavelet and discrete cosine transforms. Comput Electr Eng
7. Bayram S, Sencar HT, Memon N (2009) An efficient and robust method for detecting copy-move forgery. In: IEEE international conference on acoustics, speech and signal processing
8. Huang H, Guo W, Zhang Y (2008) Detection of copy-move forgery in digital images using SIFT algorithm. In: Pacific-Asia workshop on computational intelligence and industrial application, PACIIA'08, vol 2. IEEE, New York
9. Bo X, Junwen W, Guangjie L, Yuewei D (2009) Image copy-move forgery detection based on SURF. In: Proceedings of IEEE international conference on multimedia information network security (MINES), pp 889–892
10. Ardizzone E, Bruno A, Mazzola G (2015) Copymove forgery detection by matching triangles of keypoints. IEEE Trans Inf Forensics Secur 10(10):2084–2094
11. Yu L, Han Q, Niu X (2016) Feature point-based copy-move forgery detection: covering the non-textured areas. Multimedia Tools Appl 75(2):1159–1176
12. Pun C-M, Yuan X-C, Bi X-L (2015) Image forgery detection using adaptive over-segmentation and feature point matching. IEEE Trans Inf Forensics Secur 10(8):1705–1716
13. Christlein V, Riess C, Jordan J, Riess C, Angelopoulou E (2012) An evaluation of popular copy-move forgery detection approaches. IEEE Trans Inf Forensics Secur 7:1841–1854

Gait Recognition Using Normal Distance Map and Sparse Multilinear Laplacian Discriminant Analysis

Risil Chhatrala, Shailaja Patil and Dattatray V. Jadhav

Abstract In visual surveillance applications, gait is the preferred candidate for recognition of the identity of the subject under consideration. Gait is a behavioral biometric that has a large amount of redundancy, complex pattern distribution and very large variability, when multiple covariate exist. This demands robust representation and computationally efficient statistical processing approaches for improved performance. In this paper, a robust representation approach called Normal Distance Map and multilinear statistical discriminant analysis called Sparse Multilinear Discriminant Analysis is applied for improving robustness against covariate variation and increase recognition accuracy. Normal Distance Map captures geometry and shape of silhouettes so as to make representation robust and Sparse Multilinear Discriminant Analysis obtains projection matrices to preserve discrimination.

1 Introduction

Automated identification and recognition of the individual person in a surveillance environment with natural setting, is so hard problem that not a single gait recognition system has been reported to be working in challenging real world conditions. A large portion of the literature is dedicated to important aspects of gait recognition, so as to make it a realizable solution. The aspects like segmentation, pre-processing, gait representation schemes and pattern recognition algorithms are widely studied to understand diverse aspect requirement for gait processing.

The main contribution is as follows.

1. A new robust gait representation scheme that make use of boundary curvature information over a complete gait cycle called as Normal Distance Map is proposed.

R. Chhatrala (✉) · S. Patil
Rajarshi Sahu College of Engineering,
Savitribai Phule Pune University, Pune, Maharashtra, India
e-mail: therisil@gmail.com

D. V. Jadhav
Directorate of Technical Education, Mumbai, Maharashtra, India

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_14

2. The representation preserves the natural tensorial structure along with spatio-temporal and structural information of gait.
3. The Extracted features provide a new feature space that addresses covariates and is found to be robust for gait recognition.
4. Sparse Multilinear Laplacian Discriminant Analysis for tensor objects is used to improve discrimination capability and increase recognition rate.

This paper is structured as follows. After reviewing the literature work in Sect. 2, the gait representation scheme based on boundary curvature information over a complete gait cycle called as Normal Distance Map is presented in Sect. 3. Section 4 presents feature extraction and pattern recognition using Sparse Multilinear Laplacian Discriminant Analysis followed by experimentation in Sect. 5. Sections 6 and 7 gives Discussion and Conclusion respectively.

2 Review of Literature

Comprehensive review of the published techniques and strategies for gait as a biometric can be found in the work of Makihara et al. [1], Sivarathinabala et al. [2], Zhang et al. [3], Boulgouris et al. [4] and Wang et al. [5]. The widely researched areas of the gait recognition system is gait representation, feature dimensionality reduction and classification.

The pioneer approaches for gait descriptor are Gait Energy Image (GEI) [6], Shifted Energy Image (SEI) [7], Gait Entropy Image (GEnI) [8], Gait flow image [9], Frequency-Domain Features [10], Depth Gradient Histogram Energy Image (DGHEI) [11] and Histogram of boundary normal vector (HoNV) using local Gauss Maps [12].

3 Gait Representation

In most recent work; it is observed that, the spatio-temporal variation exhibited by gait is mainly represented by shapes and kinematic variation averaged over gait period. The work of Tang et al. [13] and El-Alfy et al. [12, 14] inspired us to use local curvatures of a silhouette contour obtained from the geometry of the silhouettes and distance transform, together to capture boundary and area information. This feature descriptor is called Normal Distance Map (NDM). In all further discussion, Normal Distance Map (NDM) as suggested by El-Alfy et al. [14] is employed as gait feature representation. Following subsection gives a brief review of work from El-Alfy et al. [12, 14].

3.1 Histogram of Boundary Normal Vector

The histograms of boundary normal vectors were introduced by El-Alfy et al. [12]. The method focuses on curvature of contours extracted from the geometry of the silhouettes. Local Gauss maps are used to link Unit vectors normal to a surface to its curvature [15, 16]. The key advantage of doing this is to come to conclusion that, "If the normal vectors magnitude is fixed and scaled to unity, then contours extracted from 'parallel' geometric curvatures need to have same histograms." [12]. This causes the descriptor to be robust to covariate that affect gait itself.

3.2 Gauss Maps

A Gauss map g is a mapping function that maps each point p from a surface $M \in R^3$ to the unit sphere S^2, with the unit vector n_p normal to M at p.

$$\begin{aligned} g : M &\rightarrow S^2 \\ p &\mapsto n_p \end{aligned} \tag{1}$$

The normal vectors to flat surface M has no variation between them and is always parallel to each other. On the other hand, variation is seen for an "overly" curved surface. Hence, the mapping g is used to model the curvature of the surface. The Gaussian curvature k is defined as:

$$k = \lim_{\Omega \to 0} \frac{\text{area of } g\,(\Omega)}{\text{area of } \Omega} \tag{2}$$

The total curvature of Ω is defined by:

$$\text{total curvature of } \Omega = \int_{\Omega} kdA \tag{3}$$

where dA is a surface element on M.

Since, Gaussian curvature of a surface $M \in R^3$ is invariant under local isometries, globally computed surface cannot always be used to discriminate. Hence, Gauss maps are defined locally, in the form of patches or cells and for each small surface or cell, local curvatures are computed.

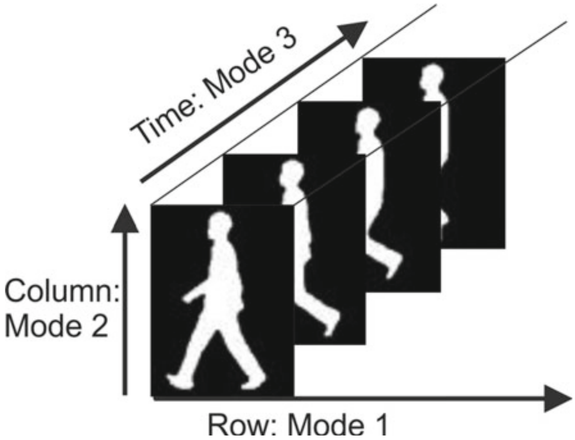


Fig. 1 Third order tensorial binary silhouette video sequence (GSV)

3.3 Normal Distance Map as Tensor

In order to compute NDM, the approach of El-Alfy et al. [14] is used. In this paper, the final feature descriptor is represented in the form of matrix for each frame. The gait cycle consists of multiple frames exhibiting inherent variation in pixel distribution due to kinematic motion, it is represented as third order binary gait silhouette volume (GSV) as shown in Fig. 1. Once, NDM for each frame from GSV is computed separately, it is repeated for all frames in GSV and as a result third order tensor representation based on NDM is obtained. It is then further processed by using statistical tensor based dimensionality reduction technique called as SMLDA [17]. Following section gives a brief overview of SMLDA.

4 Sparse Multilinear Laplacian Discriminant Analysis

It is a multilinear discriminant subspace learning derived from previous work [17]. The weighted Laplacian scatter difference along with sparsity constraint derives, sparse matrices $\mathbf{W}_k \in R^{m_k \times D_k}$, $m_k < D_k$, $k = 1, 2, 3$ that map $\mathcal{X}$ in to reduced tensor space.

$$\mathcal{Y} = \mathcal{X} \times_1 \mathbf{W}_1^T \times_2 \mathbf{W}_2^T \times_3 \mathbf{W}_3^T \in R^{m_1 \times m_2 \times m_3} \tag{4}$$

The key idea for SMLDA is to provide discrimination; when samples are from different classes and minimize variation, when samples are from the same class. It has sparsity constraints in the form of L_1 and L_2 norms penalty. The objective function is given by

$$J\left(\mathbf{W}_k\right)\Big|_{k=1}^{N} = \min tr\left(\mathbf{W}_k^T\left(\mathbf{LS}_w^{(k)} - \lambda_k * \mathbf{LS}_B^{(k)}\right)\mathbf{W}_k\right) + \alpha_k \left\|W_k\right\|^2 + \sum_j \beta_{kj}\left|w_{kj}\right| \text{ Subject to } W_k^T W_k = I_k \quad (5)$$

where $|\cdot|$ and $\|\bullet\|$denote L_1 and L_2 norm respectively, λ_k is weight parameter.

The projection matrices are computed from the gallery NDM sequence by training. Locality constrained group sparse representation (LGSR) classifier [18] is used to classify the projected low dimension features. Recognition accuracies are computed with Rank 1 ("R1") correct classification rate.

5 Experimental Evaluation

The proposed approach is validated using USF benchmark [19] and OU-ISIR data sets [20]. The recognition rate in terms of absolute as well as relative correct classification rates (CCR) are reported.

5.1 Experimentation Settings

The gait video sequence is preprocessed by background subtraction, binarization, spatial normalization, gait period detection and temporal alignment processes. As suggested in El-Alfy et al. [14], NDM is computed cell wise for each frame and the same procedure is repeated for all frames in one gait cycle. The processing for the gallery and probe sequence is repeated to obtain processed NDM tensor. Once the tensorial NDM sequence is available for both gallery and probe sets, similarity matrix based on tensor coding length is computed. The approach of minimum reconstruction error is used in the locality constrained group sparse representation (LGSR) is used for classification of probe binary gait tensor in to one from a gallery.

5.2 Experiments on OU-ISIR Dataset

OU-ISIR is the worlds largest and widely preferred available gait database with 1872 females and 2135 males totaling more than 4007 subjects with maximum diversity of covariate.

Experiment on OU-ISIR dataset A As suggested by El-Alfy et al. [12] all guidelines for experimental protocol are followed. The entire database is divided in to five sets as A-55 to A-85 and A-All. Table 1 shows the performance of the proposed technique.

Table 1 Recognition rate for OU-ISIR dataset

Dataset	No. of Subjects	GEI	HONV	NDM [14]	NDM SMLDA (TCL)
A-55	3706	84.7	91.6	94.1	95
A-65	3770	86.6	92.1	95	96.2
A-75	3751	86.9	93.3	95.7	97
A-85	3249	85.9	93.0	95.9	96.6
A-All	3141	94.2	97.5	98.1	97.2
Average		87.6	93.5	95.8	96.4

5.3 *Experiment on USF Database*

The USF dataset by Sarkar et al. [19] has a total of 1870 gait sequences captured in an outdoor environment for 122 subjects. Entire dataset is divided into two categories, one for the gallery and another category as twelve probe sets (A–L) having diverse variations: like shoe, surface, view points, carrying condition and time. Table 2 shows the performance of proposed technique.

Table 2 Rank-1 correct classification rate (%) for USF dataset

Probe set	Probe size	Baseline [19]	CGI Fusion [21]	Gabor-PDF + LGSR [18]	Gabor + RSM + HDF [22]	NDM + SMLDA (TCL)
A	122	73	91	95	100	100
B	54	78	93	93	95	97
C	54	48	78	89	94	91
D	121	32	51	62	73	78
E	60	22	53	62	73	75
F	121	17	35	39	55	60
G	60	17	38	38	64	50
H	120	61	84	94	97	87
I	60	57	78	91	99	93
J	120	36	64	78	94	93
K	33	3	3	21	42	45
L	33	3	9	21	42	45
Average		43	56	65	77	76

6 Discussion

Comparing the results in Tables 1 and 2, the following observations can be drawn:

1. When single-template-based or matrix based gait representation approach is compared with the third-order tensor representation, it outperforms and the average correct recognition rate is much improved. The simplest justification is the preservation of inherent structural information.
2. The covariate leads to partial feature corruption problems. The proposed representation preserves inherent correlation by retaining tensor based structural information. Simply by avoiding vectorization, robustness are improved by restricting corruption of features.
3. Tensor coding length (TCL) approach uses natural correlation of pixels from their spatial locations to reduce contamination thereon.

6.1 Timing Analysis

The timing analysis of the discriminative feature extraction scheme and the iterative projection-method-based optimization procedure of SMLDA is done by measuring the amount of time required for execution. The code of our method is run on a PC with an Intel Core i7 3.5 GHz processor and 16GB RAM. For USF dataset, the training time for the tensor coding length approach is reported as 300 s and query time as 0.42 s.

7 Conclusion

In this paper, a robust representation approach called Normal Distance Map and multilinear statistical discriminant analysis called Sparse Multilinear Discriminant Analysis is applied for improving robustness against covariate variation and increase recognition accuracy. NDM captures geometry and shape of silhouettes so as to make representation robust and SMLDA obtains projection matrices that preserve discrimination.

References

1. Makihara Y, Matovski DS, Nixon MS, Carter JN, Yagi Y (2015) Gait recognition: databases, representations, and applications. Wiley Online Library
2. Sivarathinabala M, Abirami S, Baskaran R (2017) A study on security and surveillance system using gait recognition. In: Intelligent techniques in signal processing for multimedia security. Springer, Berlin, pp 227–252

3. Zhang Z, Hu M, Wang Y (2011) A survey of advances in biometric gait recognition. In: Chinese conference on biometric recognition. Springer, Berlin, pp 150–158
4. Boulgouris NV, Hatzinakos D, Plataniotis KN (2005) Gait recognition: a challenging signal processing technology for biometric identification. IEEE Signal Process Mag 22(6):78–90
5. Wang J, She M, Nahavandi S, Kouzani A (2010) A review of vision-based gait recognition methods for human identification. Digit Image Comput: Tech Appl pp 320–327
6. Han J, Bhanu B (2006) Individual recognition using gait energy image. IEEE Trans Pattern Anal Mach Intell 28:316–322
7. Huang X, Boulgouris NV (2012) Gait recognition with shifted energy image and structural feature extraction. IEEE Trans Image Process 21:2256–2268
8. Bashir K, Xiang T, Gong S (2009) Gait recognition using gait entropy image. In: In 3rd international conference on crime detection and protection, London, UK
9. Lam THW, Cheung K, Liu JN (2011) Gait flow image: a silhouette-based gait representation for human identification. Pattern Recogn 44:973–987
10. Makihara Y, Sagawa R, Mukaigawa Y, Echigo T, Yagi Y (2006) Gait recognition using a view transformation model in the frequency domain. In: European conference on computer vision. Springer, Berlin, pp 151–163
11. Hofmann M, Bachmann S, Rigoll G (2012) 2.5 d gait biometrics using the depth gradient histogram energy image. In: 2012 IEEE fifth international conference on biometrics: theory, applications and systems (BTAS). IEEE, New York, pp 399–403
12. El-Alfy H, Mitsugami I, Yagi Y (2014) A new gait-based identification method using local gauss maps. In: Asian conference on computer vision. Springer, Berlin, pp 3–18
13. Tang S, Wang X, Lv X, Han TX, Keller J, He Z, Skubic M, Lao S (2012) Histogram of oriented normal vectors for object recognition with a depth sensor. In: Asian conference on computer vision. Springer, Berlin, pp 525–538
14. El-Alfy H, Mitsugami I, Yagi Y (2017) Gait recognition based on normal distance maps. IEEE Trans Cybern
15. Gauss KF (1902) General investigations of curved surfaces of 1827 and 1825
16. Hazewinkel M (2001) Encyclopaedia of mathematics, vol 13. Springer, Berlin
17. Chhatrala R, Patil S, Lahudkar S, Jadhav DV (2017) Sparse multilinear Laplacian discriminant analysis for gait recognition. Pattern Anal Appl pp 1–14
18. Xu D, Huang Y, Zeng Z, Xu X (2012) Human gait recognition using patch distribution feature and locality-constrained group sparse representation. IEEE Trans Image Process 21(1):316–326
19. Sarkar S, Phillips P, Liu Z, Vega IR, Grother P, Bowyer K (2005) The humanid gait challenge problem: data sets, performance, and analysis. IEEE Trans Pattern Anal Mach Intell 27:166–177
20. Iwama H, Okumura M, Makihara Y, Yagi Y (2012) The ou-isir gait database comprising the large population dataset and performance evaluation of gait recognition. IEEE Trans Inf Forensics Secur 7(5):1511–1521
21. Wang C, Zhang J, Pu J, Yuan X, Wang L (2010) Chrono-gait image: a novel temporal template for gait recognition. In: European conference on computer vision. Springer, Berlin, pp 257–270
22. Guan Y, Li CT, Roli F (2015) On reducing the effect of covariate factors in gait recognition: a classifier ensemble method. IEEE Trans Pattern Anal Mach Intell 37(7):1521–1528

Comparative Study and Analysis of Pulse Rate Measurement by Vowel Speech and EVM

Ria Paul, Rahul Shandilya and R. K. Sharma

Abstract The paper presents two noncontact pulse rate measurement techniques from vowel speech signals and Eulerian Video Magnification. The proposed methods use signals those are neither audible nor visible to naked eyes. The signals are recorded and their characteristic plots and spectrum analysis by Short-Time Fourier Transform reveal some peaks from which pulse rate can be calculated. The methods are then compared with the conventional methods where the accuracy differs by only 3.9% for vowel speech and by 0.4% for Eulerian Video Magnification. The Bland–Altman plot for the techniques shows that both are acceptable as they lie between ±1.96 Standard Deviation. The data collected from the methods are processed in MATLAB and also implemented on FPGA using serial communication by RS232.

1 Introduction

Heart rate is the number of times of contraction and expansion of heart per minute. Heart rate depends on many factors. A normal resting heart beats at 60–100 Beats Per Minute (BPM). Heart rate depends on the body's need to absorb oxygen and therefore varies during physical activity, various emotional stage, fitness, medications, temperature, etc. Heart rate is traditionally measured by counting arterial pulsation by fingers. The heart activity can be detected by electrocardiogram using 12 lead electrodes. The oxygen saturation of blood and pulse rate is measured from com-

R. Paul (✉) · R. Shandilya · R. K. Sharma
School of VLSI Design and Embedded Systems, National Institute of Technology Kurukshetra, Kurukshetra, Haryana, India
e-mail: ria.paul1993@gmail.com

R. Shandilya
e-mail: rss.nitk@gmail.com

R. K. Sharma
e-mail: mail2drrks@gmail.com

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_15

mercial oximeters. These conventional technologies are significant in monitoring of cardiac patients.

The conventional heart rate monitoring methods are time-consuming and costly. The patient needs to visit the clinic for checkup. In times of emergency when the condition of the patient is very critical and there is no clinic nearby, the conventional methods become inconvenient. Also, for a cardiac patient who cannot afford regular heart checkup, proposed the methods which become very convenient as it requires only a microphone or a camera. This paper discusses about design and comparative analysis of two noninvasive methods by which one can easily detect pulse rate and heart condition without visiting a clinic. The first method is vowel speech [1, 2] signal processing recorded by using standard microphones. The second method is Eulerian Video Magnification (EVM) [3, 4] by video and image processing of red-plane intensity of video streaming from finger tip by passing LED light through it.

The proposed methods are then implemented on FPGA by sending collected data via RS232 from MATLAB and storing them on BRAM. The stored data are then analyzed and pulse rate is detected. Both the methods require the same architecture. The only difference is that for the first method data is collected by recording of vowel such as "e" by a standard microphone and the second method requires a camera and a LED. All the components are easily available, affordable, and convenient. The accuracy of the second method that is Eulerian Video Magnification is very high and can be easily used in times of emergency.

2 Architecture

2.1 Pulse Rate Measurement Using Vowel Speech Signals

Human speech signals contain biological and semantic information. The speech outputs are results of activities of sound energy source like larynx and supralaryngeal vocal tract. According to the "source-filter theory of speech production", speech output is the result of two-stage process (i) generation of sound energy with some spectrum and (ii) filtration or modification of source sound by vocal tract. The larynx has muscles which contain blood vessels and as part of the cardiovascular system, these vessels have connection to the heart. Heartbeat is also related with length and volume of vocal cord. Thus, heartbeat can easily be detected from human speech. In our experiment, we record the vowel "e" and plot the spectrum for heartbeat analysis. Recording rooms require high sound quality and less noise. Walls of porous materials like sponges, wood, and melamine create less noise than concrete walls. The pulse rate is detected from the spectrum by the given steps.

Steps to measure pulse rate from fingertip are shown in Fig. 1 as:

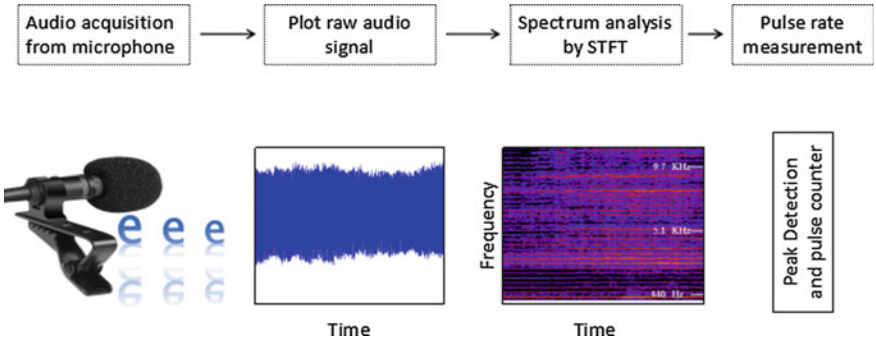


Fig. 1 Block diagram of pulse rate measurement from vowel speech recording

Step 1—The speech of any vowel is recorded using a microphone. The microphone of a mobile phone can also be used. The vowel recorded in this experiment is "e".

Step 2—The plot of the audio signal is done with respect to time but this does not give any information about frequency. The graph is again analyzed for obtaining frequency components.

Step 3—The speech signal is nonstationary and changes over time. If we apply DFT or FFT over long window, no information is obtained about frequency. Therefore, Fourier transform is applied over short period of time because over short duration or short window, speech signal can be considered stationary. The size of the window used should be less than the R-R interval of heartbeat. The windows must be tapering like Hann, Hamming and also overlapping. The technique used is called Short-Time Fourier Transform (STFT) [5, 6]. The STFT also filters unnecessary noises from the speech signal. The total number of peaks over total gives pulse rate per second. The pulse rate per second is multiplied by 60 to get pulse rate or heart Beats Per Minute (BPM). Equation (1) gives formula for STFT where $x[n]$ is the signal and $w[n]$ represents the window used.

$$X(n, \omega_0) = \sum_{m=-\infty}^{\infty} \left(x[m]\mathrm{e}^{-j\omega_0 m}\right) w[n-m] \tag{1}$$

2.2 Pulse Rate Measurement Using Eulerian Video Magnification

By using camera, the proposed approach was implemented. This camera is used to absorb the light from blood samples. That oxygenated blood will move to the next step. Based on the intensity of the light, it can estimate the volume of blood. Based on the following steps, pulse rate can be absorbed.

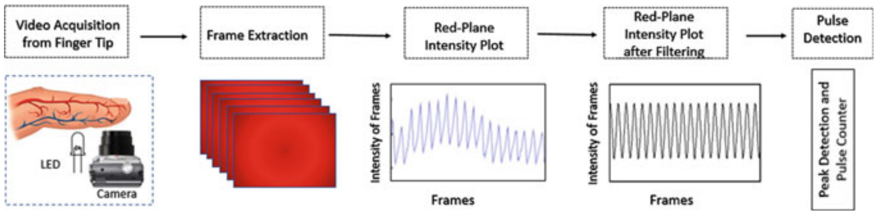


Fig. 2 Block diagram of pulse rate measurement from EVM

Fig. 3 Block diagram for FPGA implementation of the techniques

Steps to measure pulse rate from fingertip are shown in Fig. 2 as:

Step 1—By using the camera, it captures the fingerprint and it will produce the video that will be frames per second and it is very minute. This process is called Eulerian Video Magnification (EVM) [7] and is applied to fingertip of our hands or earlobe to measure pulse rate. These regions are chosen because they do not have any bones to block the light and measurement can be done accurately.

Step 2—Frame extraction process will be done through each frame of the video.

Step 3—The high-pass filters remove the low-frequency variations including non-linearity and low-pass filters remove the high-frequency noises. The plot is further made linear by using detrend( ) function in MATLAB. Now, the peaks are detected from the filtered data. The total number of peaks over total time of video streaming gives pulse rate per second. The pulse rate per second is multiplied by 60 to get pulse rate or heart Beats Per Minute (bpm).

2.3 Design for Implementation of the Methods on FPGA

The audio or video signal is stored in MATLAB which sends the data to FPGA via RS232 by serial communication [8]. The BRAMs on the FPGA are programmed to store the data. The FPGA then processes the data from BRAM and after detecting the peaks from the given data, the pulse rate per minute is measured.

Steps to measure pulse rate by FPGA implementation are shown in Fig. 3 as:

Step 1—The audio or video signal required for the heartbeat analysis are obtained from microphone or camera and are sent for MATLAB analysis in PC. The raw data

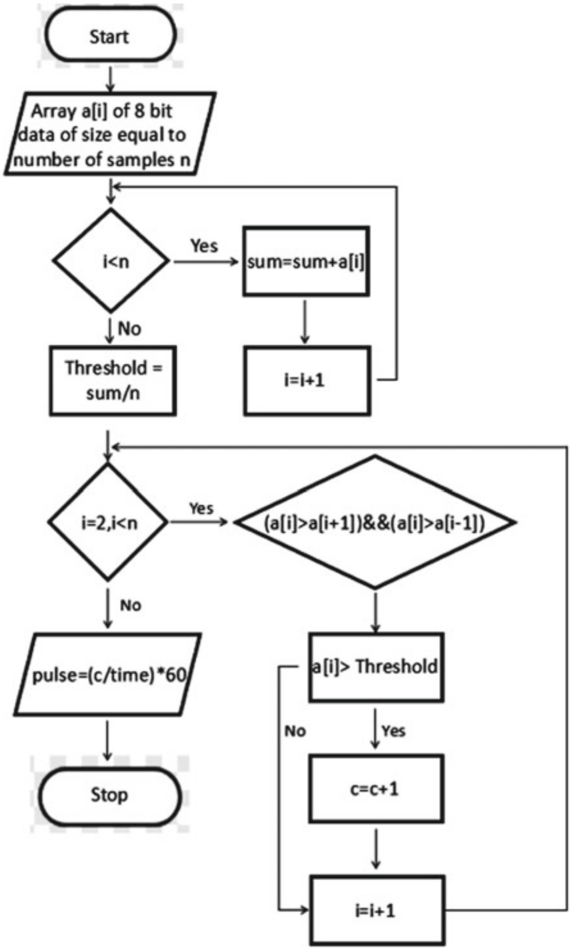


Fig. 4 Flow chart for pulse rate measurement from peak detection in Verilog

is analyzed by MATLAB where noises and false peaks are removed. The filtered data are now sent to FPGA using serial communication by RS232.

Step 2—The BRAM on FPGA are programmed to store the received data. The stored data are then processed by FPGA to find out the peaks in the data. The flowchart for peak detection and pulse rate measurement using Verilog on FPGA is shown in Fig. 4.

Step 3—The pulse rate is shown on LCD display of FPGA and also sent back to PC for display.

3 Results and Discussion

The techniques are processed in MATLAB and then implemented on FPGA Spartan 3E kit. The microphone used is Beyer dynamic 716413 MMX 102iE with sensitivity of 104 dB.

3.1 Pulse Rate Measurement by Vowel Speech Recording

The raw data signal is plotted first and then the spectrum analysis is done using Short-Time Fourier Transform which is further analyzed for peak detection and pulse rate measurement. The recording of vowel "e" is done for 10 s and the plot of the signal versus time is shown in Fig. 5. The plot does not give any direct information about pulse rate. Therefore we go for spectrum analysis of the audio signal.

The spectrum analysis done by Short-Time Fourier Transform in Fig. 6 shows the peaks and it can be further analyzed to show the peaks distinctively as samples. The total number of peaks over total time gives the pulse rate per second. Equation (2) gives pulse per minute.

$$\text{Pulse Rate} = \frac{\text{Number of peaks detected}}{\text{Duration of vowel speech}} \times 60 \tag{2}$$

3.2 Pulse Rate Measurement by Eulerian Video Magnification

300 frames of video stream are acquired. The red-plane intensity of each frame of the video is plotted in Fig. 7, and the graph clearly shows the peaks of intensity variation that represents heartbeat with time. The plot is then filtered to remove false peaks and it is also made linear.

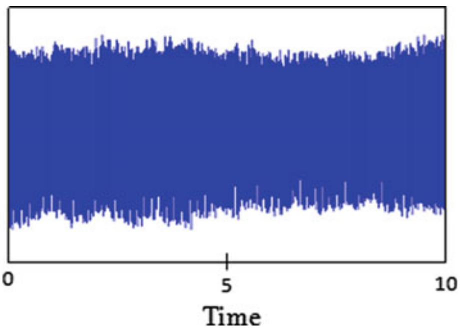


Fig. 5 Vowel speech signal versus time

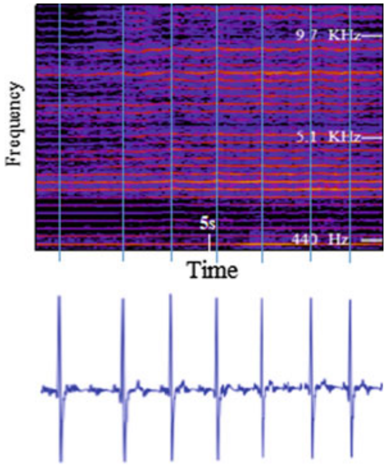


Fig. 6 STFT of vowel speech signal for spectrum analysis

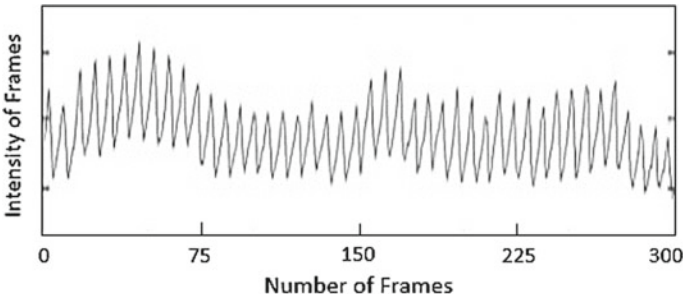


Fig. 7 Red-plane intensity graph of video stream captured from fingertip

The plot in Fig. 8 shows the intensity graph after passing through the low-pass and high-pass filters and removing nonlinearity.

The pulse per minute is given by Eq. (3). The pulse rate using commercial oximeter, vowel speech technique and Eulerian Video Magnification for 10 subjects are compared in Table 1. The average accuracy of the proposed vowel speech technique is 96.1% and average accuracy of proposed Eulerian Video Magnification technique is 99.6%. Therefore, Eulerian Video Magnification is more accurate and can be easily used as it has 99.6% accuracy. The vowel speech method is also acceptable as it has more than 95% accuracy.

$$\text{Pulse Rate} = \frac{\text{Number of peaks detected}}{\text{Duration of video stream}} \times 60 \quad (3)$$

Figure 9 shows Bland–Altman plot [9, 10] of ten subjects for the above two techniques. Bland–Altman plot is used to compare the agreement between two methods.

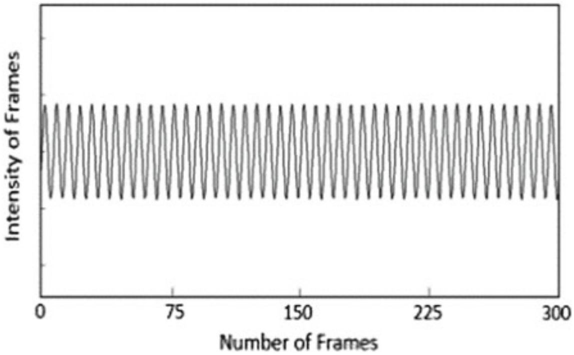


Fig. 8 Red-plane intensity graph after filtering

Table 1 Comparison of pulse rate recorded by proposed techniques

Subject	Age	Weight	Pulse rate (/min) Using commercial oximeter (A)	Pulse rate (/min) Using vowel speech (B)	Error (%) in vowel speech ((A-B)/A) × 100	Pulse rate (/min) Using EVM (C)	Error (%) in EVM ((A-C)/A) × 100
1	23	50	77	80	3.9	78	1.3
2	24	48	89	87	2.25	89	0
3	35	75	69	71	2.9	69	0
4	21	80	87	90	3.45	86	1.15
5	45	85	83	80	3.6	83	0
6	25	70	64	61	4.69	63	1.56
7	22	45	91	87	4.4	92	1.1
8	23	58	72	76	5.55	73	1.39
9	25	55	80	76	5	80	0
10	24	60	92	89	3.26	92	0

The plot shows the difference in parameters versus average of the parameters. The plot shows pulse rate of each of the ten subjects by vowel speech and Eulerian Video Magnification. It is observed that all the values fall within ±1.96 Standard Deviation which means the methods have a good agreement and thus both the methods can be used for pulse rate measurement.

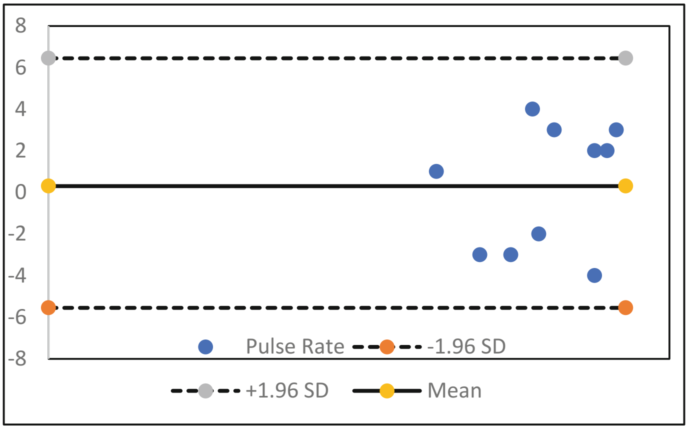


Fig. 9 Bland–Altman plot for pulse rate measurement by vowel speech and EVM

4 Conclusion

The noncontact pulse rate measurement techniques are especially suited in cases of emergency for heart patients. They also help to keep a track of the heart's condition for cardiac patients at no cost. The vowel speech method is less reliable compared to traditional oximeters but can be easily used as they have 96.1% accuracy. The Eulerian Video Magnification technique has 99.6% accuracy and is easily reliable and authentic. The techniques discussed in the paper only need a smartphone as it has both LEDs and camera which is needed for EVM technique. The vowel speech technique needs an earphone with speakers which are very commonly available. Thus, the techniques require zero cost. The vowel speech technique can also be used to analyze different emotional stages [11]. However, these methods are not applicable for patients with artificial heart and pacemakers. The successful implementation of the techniques on FPGA assures that in future a device can be designed for pulse rate measurement and heartbeat analysis using vowel speech recording and Eulerian Video Magnification.

References

1. James AP (2015) Heart rate monitoring using human speech spectral features. Hum-Centric Comput Inf Sci 5(1):33
2. Mesleh A et al (2012) Heart rate extraction from vowel speech signals. J Comput Sci Technol 27(6):1243–1251
3. Sukaphat S et al (2016) Heart rate measurement on Android platform. In: 2016 13th international conference on electrical engineering/electronics, computer, telecommunications and information technology (ECTI-CON). IEEE, New York, pp 1–5

4. He X, Goubran RA, Liu XP (2016) Wrist pulse measurement and analysis using Eulerian video magnification. In: 2016 IEEE EMBS international conference on biomedical and health informatics (BHI). IEEE, New York, pp 1–4
5. Aubel C, Stotz D, Bölcskei H (2018) A theory of super-resolution from short-time Fourier transform measurements. J Fourier Anal Appl 24(1):45–107
6. Zhang G (2018) Time-phase amplitude spectra based on a modified short-time Fourier transform. Geophys Prospect 66(1):34–46
7. He X, Goubran RA, Liu XP (2014) Using Eulerian video magnification framework to measure pulse transit time. In: 2014 IEEE international symposium on medical measurements and applications (MeMeA). IEEE, New York, pp 1–4
8. Pokharkar S (2015) FPGA based design and implementation of ECG feature extraction. Int J Adv Found Res Sci Eng (IJAFRSE) 1(12)
9. Altman DG, Bland JM (2017) Assessing agreement between methods of measurement. Clin Chem 63(10):1653–1654
10. Matsuura N et al (2017) Bland–Altman analysis for method comparisons. Adv Mod Med 354
11. Ryskaliyev A, Askaruly S, James AP (2016) Speech signal analysis for the estimation of heart rates under different emotional states. In: 2016 International conference on advances in computing, communications and informatics (ICACCI). IEEE, New York

Real-Time Input Text Recognition System for the Aid of Visually Impaired

B. K. RajithKumar, H. S. Mohana, Divya A. Jamakhandi, K. V. Akshatha, Disha B. Hegde and Amisha Singh

Abstract It is estimated that 285 million people globally are visually impaired. A majority of these people live in developing countries and are among the elderly population. Reading is essential in daily life for everyone. Visually impaired persons can read only by use of special scripts specially designed for them such as Braille language. Further, only trained people can read and understand. Since every product does not provide the product information on product cover in Braille, the present work proposes an assistive text reading framework to help visually impaired persons to read texts from various products/objects in their daily lives. The first step in implementation captures the image of the required by extracting frames from real-time video input from the camera. This is followed by preprocessing steps which includes conversion to grey scale and filtering. The text regions are further extracted using MSER followed by canny edge detection. The text regions from the captured image are then extracted and recognized by using Optical Character Recognition software (OCR). The OCR engine Tesseract is used here. This extracts the text of various fonts and then sizes can be recognized individually and then combined to form a word. Further, producing audio output by using Text to Speech module. The result obtained is very much comparable with other existing methods with better time efficiency. The real-time input is taken and passed through the algorithm which applies filters and removes noise then later image is passed through MSER, OCR, Canny edge detection to get the final audio output.

B. K. RajithKumar (✉) · D. A. Jamakhandi · K. V. Akshatha · D. B. Hegde · A. Singh
Department of Electronics and Communication Engineering, R V College of Engineering, Bengaluru, India
e-mail: Rajith.bkr@rvce.edu.in

K. V. Akshatha
e-mail: Akshatahakv.ec16@rvce.edu.in

H. S. Mohana
Department of Electronics and Instrumentation Engineering, Malnad College of Engineering, Hassan, India
e-mail: Hsm@mcehassan.ac.in

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_16

1 Introduction

Although there are many, but not all things/products which have the text printed in a manner which an average blind person can recognize [1]. While a lot of visually impaired seek assistance of their fellow pedestrians in acquiring data from public information boards, the need of the hour is the availability of cost-effective, robust and accurate module which can identify the text from an image captured and give the output in a way that can be utilized comfortably by a visually impaired person, which makes them confident and independent enough so as they make their own way in this fast-growing society. We have thus developed a module which can take images from a real-time video camera, identify text, and give the output as an audio file so as to aid the visually impaired person with the information on public information boards.

The problem in achieving the task is to automatically localize and detect the text from real-time captured images with different backgrounds [2]. The existing Optical Character Recognition (OCR) technique can identify text from a clean picture of a sheet, the challenge in the current situation is to recognize text from the images with extreme noise, text with various fonts and sizes. This has been tackled by adopting various image filtering techniques. These techniques involve Canny edge detection, maximally stable extremal regions, and OCR. These techniques help us to eliminate noise in every stage. The images from the camera are captured in the form of frames and are processed real time to extract text and convert the obtained text to speech signals.

2 Related Works

Most text detection algorithms utilize a similar flow of operations to extract the regions of interest, followed by the classification of the text. These operations or methods include the identification of Maximally Stable Extremal Regions (MSER), Canny Edge Detection, and some form of OCR [3–5]. These have been determined to increase the efficiency of test recognition as compared to other methods when using novel OCR techniques [4]. There are three existing methods of text recognition: (1) texture-based method [6]; (2) Connected Component (CC) based method [7, 8]; and (3) hybrid method [9]. In texture based method the text region is considered to have a different distinguishable texture from the background. A classifier can be trained to identify the existence of text on the image. If images are more complicated, texture segmentation scheme is not sufficient. Connected component based method carries out connected component analysis and then the grouping of character candidates into text. Noncharacters are erased by Conditional Random Fields (CRFs).

First, MSER approaches are applied to a scene image to detect a large region of noncharacter which employs MSER as a basic Connected Component. The implementation of MSER may be carried out by first performing a luminance threshholding on the image in question. The extracted regions give rise to various grouping hypotheses, which can be narrowed down by examining the properties of different regions using region filtering. The paper [3] has made use of region filtering to identify the properties of different regions using pixel values. These properties can be used to separate the image into sub-images. The relationship of regions with respect to colour similarity, shape similarity can also be considered [5]. To enhance the performance, MSER must be combined with canny edge detection, which outlines the text characters and can be optimized by varying threshholding values. A Canny edge is the weak edge of local maxima of the gradient of the natural image. The noisy regions of the image can be filtered out by performing reliable geometric checks on the connected components [4]. OCR is one of the most successful methods of character recognition [3]. The entire OCR is mainly classified into two categories: traditional OCR and object recognition based. For traditional OCR-based methods, various binarization methods have been proposed to get the binary image which is directly fed into the off-the-shelf OCR engine. On the other hand, object recognition based methods assume that scene character recognition is quite similar to object recognition with a high degree of intra-class variation for scene character recognition; these methods directly extract features from original images and use various classifiers to recognize the character. OCR engines output in the form of text. In conclusion, the overall benefit of combining the methods of MSER, Canny Edge detection and OCR are evident in the cleaner output obtained. In [5, 10], their novel OCR technique obtained a precision of 89.3% and was 77.47% successful in f measure. Overall studies show that the combined usage of these methods leads to a cleaner image that is ideal for OCR and conversion into text. In this proposed work, similar flow was used, by implementing MSER techniques followed by canny edge detection, as these papers have shown the benefit of pairing these methods together. Hence, an OCR engine for character recognition was used.

3 Algorithm

Step 1: Real-time video recording is done. Frames are captured from this video to process and extract text.
Step 2: The captured 3D image is then converted into a 2D grey scale image.
Step 3: Bilateral filter is applied to remove the background noise from the image.
Step 4: MSER is applied to get the sharp edges like text regions change in pixel intensities.
Step 5: Canny edge detection is done to detect all the sharp edge boundaries mainly to extract the characters from the image.
Step 6: Dilation and intensification are done to fill in the above-drawn edges.
Step 7: This image is given as an input to the OCR engine which gives an array of characters present in the picture.
Step 8: This string is then passed to the TTS engine that converts text to speech and this audio is then fed into the earphones to help the disabled.

4 Methodology and Implementation

The block diagram of the proposed module is shown in Fig. 1.

4.1 Image Conversion

Processing a 2D array is simpler than processing a 3D array, the captured frames have three channels red, green and blue. So to convert this to grey scale (2D array) standard NTSC conversion, this calculates the effective luminance of each pixel. Instead, the Ibgr2color function of open CV can also be used shown in Figs. 2 and 3.

$$I_{gray} = 0.2989 * I_{rgb}(, 1) + 0.5870 * I_{rgb}(, , 2) + 0.1140 * I_{rgb}(, , 3) \quad (1)$$

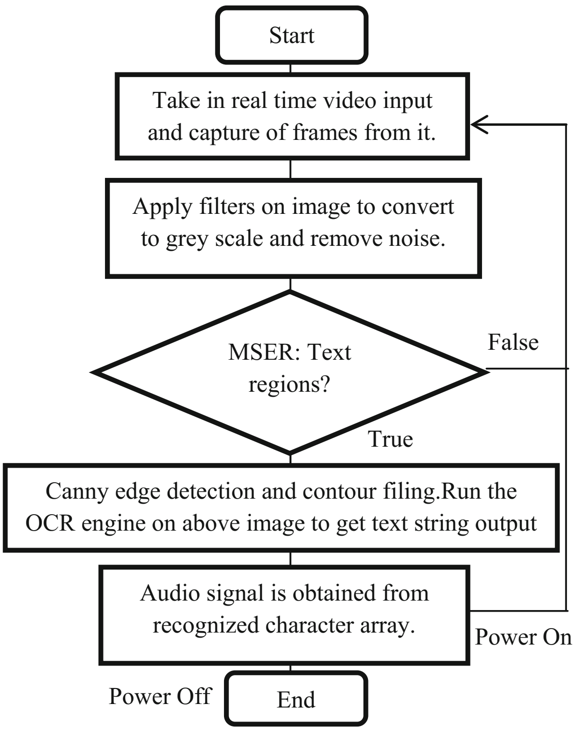


Fig. 1 Block diagram of the proposed system

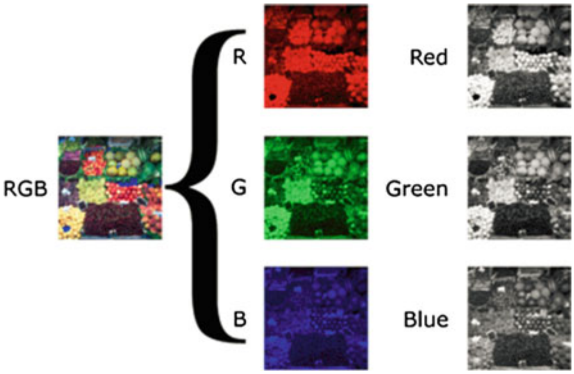


Fig. 2 Colour (3D) to grey scale conversion (2D)

Fig. 3 MSER applied image

4.2 Bilateral Filter

A bilateral filter is an image smoothing and noise removal tool which replaces the intensity of each pixel with a weighted average of intensity values from nearby pixels. Gaussian distribution is the principles behind this weighted average. These weighted averages depend on both the Euclidean distance of pixels and on the radiometric differences, therefore preserving sharp edges. The following filters are equations are given below in (2) and (3).

The bilateral filter can be defined as

$$I^{\text{filtered}}(x) = \frac{1}{W_p} \sum_{x_i \in \Omega} I(x_i) f_r(\|I(x_i) - I(x)\|) g_s(\|x_i - x\|), \tag{2}$$

where the normalized term is

$$W_p = \sum_{x_i \in \Omega} f_r(\|I(x_i) - I(x)\|) g_s(\|x_i - x\|) \tag{3}$$

4.3 MSER

On these images, MSER is applied to detect the regions of text and the images without text regions are discarded to reduce further processing on such images.

MSER is a method for blob detection in the images; it is a stable connected component of some grey level sets of the image. MSER depends on the threshold of the image, if we give them some threshold value the pixels below that threshold value are 'white' and all those above or equal are black. MSER detect the objects and all the objects can be filled with different colours in this process some of the regions include the extra background pixels.

Mathematically, let $Q1,\ldots, Qi - 1, Qi,\ldots$ be a series of nested extremal regions (Qi, $Qi + 1$). Extremal region $Qi*$ is maximally stable if and only if $q(i) = |Qi +$ ¨\ $Qi -$ ¨$|/|Qi|$ has a local minimum at $i*$. (Here | | denotes cardinality). The equation checks for regions that remain stable over a certain number of thresholds. If a region $Qi+$¨ is not significantly larger than a region $Qi-$¨, region Qi is taken as a maximally stable region. It proceeds by first sorting the pixels by intensity. After sorting, pixels are marked in the image, and the list of growing and merging connected components and their areas is maintained using the union-find algorithm in practice these steps are very fast.

Those are removed in the Canny edge detection process.

The procedure of implementation of MSER includes the following four steps:

i. Simple luminance threshholding: First of all sweep threshold of intensity from black to white performing a simple luminance threshholding of the image.
ii. Then the connected components are extracted ('Extremal Regions').
iii. A threshold is found when an extremal region is maximally stable.
iv. Finally, the regions descriptors as features of MSER are received.

4.4 Canny Edge Detection

The selected images go through canny edge detection is applied which marks the sharp boundaries of the region thus drawing an outline for the text. Canny edge detection is a tool used to extract sharp contours that is intense pixel densities, or regions with a high gradient from the background objects that do not have sharp variations. This marks these sharp variations in pixel thus reducing the amount of data to be processed on the image.

Canny edge processing has mainly five steps:

i. Noise is removed by applying a Gaussian filter to the image.
ii. Pixel intensity and the gradients of all regions in the image is found.
iii. Spurious response to edge detection is eliminated by non-maximum separation.
iv. Potential edges are determined by double thresh holding.

v. All weak edges and edges not connected to strong edges are suppressed to get the final edge marking; this is called edge marking using hysteresis.

Image noise easily affects edge detection; therefore it is necessary to filter out the noise to prevent the wrong detection caused due to noise. This filter slightly removes noise before edge detection. The filter kernel of size $(2k + 1) * (2k + 1)$ is as follows.

$$H_{ij} = \frac{1}{2\pi\sigma^2}\exp\left(-\frac{(i-(k+1))^2 + (j-(k+1))^2}{2\sigma^2}\right); \quad 1 \le i,\ j \le (2k+1) \quad (4)$$

The intersection of the Canny edge and the MSER is most likely to be a region containing text. These edges and detected text region are dilated to complete open boundaries. These completed boundaries are then filled in this step intensifies the text and thus it can be sent as input to the OCR engine.

4.5 Optical Character Recognition

OCR is an engine which has a huge library of all possible characters and using images that have discrete text characters involved in them it gives a string or character array as output. The intensified images are then passed through the OCR engine which gives the text from the image. OCR requires pre-processing which is done by the above filters like greyscale filter, bilateral filter, MSER, and canny edge. There are two basic types of core OCR algorithm, which may produce a ranked list of candidate characters. Matrix matching involves comparing an image to a stored character (glyph) on a pixel-by-pixel basis. This relies on the input glyph being correctly isolated from the rest of the image, and on the stored glyph being in a similar font and on the same scale. This technique could be better implemented with a predefined set of fonts than on a complete never style of characters. Feature extraction is also a type of OCR implementation which decomposes glyphs into "features" like lines, closed loops, line direction, and line intersections. The recognition process is made computationally efficient by reducing the dimensionality by extracting features. The extracted features are compared with a vector-type representation which may also reduce one or more glyph prototypes.

4.6 Text to Speech Engine

The text output from the OCR engine is then passed to text to speech conversion engine 'TTS' which converts this to the speech signal and this is given as output to the speaker. TTS is a python library which we use to convert text to speech. Any text input is converted into speech with default language as English, further using more libraries other language output can also be developed.

5 Results and Discussions

The following algorithm works with an efficiency of 90% (after 15 trials). Some bright strains of light during video capture can disrupt the functioning by erroneously detecting some text characters. The below figures are the output of the algorithm after each stage of image processing and also the final text output of module that gets converted to speech.

Figure 4 shows the frame captured from the real-time video input. The value of each pixel is a single sample representing only an amount of light, that is, it carries only intensity information for processing; this is shown in Fig. 5. Figure 6 shows the application of bilateral filter which sharpens the text and removes noise. Then MSER is applied which is depicted in Fig. 7. This method used for blob detection in images. It extracts a comprehensive number of corresponding image elements contributes to the wide-baseline matching. Canny edge detection is performed on the MSER output to get the output as shown in Fig. 8. Further dilation and intensification filters are applied to get the input to the OCR engine as shown in Figs. 9 and 10. Figure 11 shows the string output of the OCR engine which is fed to the Text to Speech Engine.

Fig. 4 Frame from the captured real-time video

Fig. 5 Greyscaled image

Fig. 6 Bilateral filter application

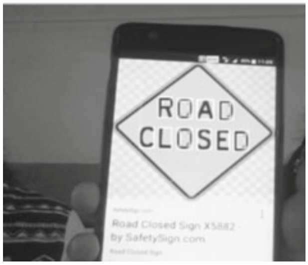


Fig. 7 MSER application

Fig. 8 Canny edge detection

Fig. 9 Dilation

Fig. 10 Intensified image (input to OCR)

5.1 Efficiency Based on Varying Parameters

Distance: Efficiency was observed to decrease by a small amount with increasing distance. It is highly dependent on the resolution of the camera or imaging device as more blur is known to reduce the accuracy of MSER technique. The efficiency reduced from 95% at 5 m to 90% at 20 m to 80% at 40 m distance.

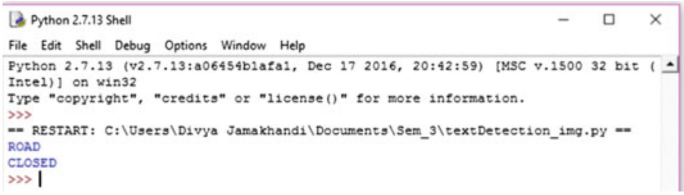


Fig. 11 Output of OCR (input to TTS)

5.2 Brightness Levels

The ambient light in the captured image varies the output due to improper edge detection. Concentrated bright spots in the image were found to obscure the edges of characters in uneven light settings. This affected similarly shaped characters more significantly than others, i.e., 'c' and 'o', 'f' and 't', 'P' and 'R', etc., due to partial obscurement of the characters. Efficiency was found to be maximum in natural light with a decrease in settings with an artificial source pointed directly or indirectly on the sign. Low light setting also brought a decrease in efficiency, but this was also dependent on the camera used.

6 Conclusion

The proposed module uses a camera, embedded on spectacles which serve as visual assistance for visually impaired. With the proposed module, visually impaired would be able to commute easily from one place to another, independently without seeking much assistance. Despite a small reduction in accuracy with increasing distance, it shall not affect the effectiveness of the module in real-world situations. The efficiency was seen to be maximum in natural light, which is the normal use-case for such a module. The future expansion would be implementing a GPS module in the system so that if the user feeds in the destination, using voice recognition it would detect the place and give commands to the destined places. With various kinds of inputs being given, considerable accuracy was achieved during the testing phase. The sequence of elimination of non-text frames and noise elimination prove to be a good choice in real-time text recognition system. The text to speech conversion library that was used gives appreciable speech output from the text output of numerous levels of filters of text recognition system. Overall, the approach of providing an aid to visually impaired through a technological product which can read out the essential text on the public information boards would help millions of visually impaired to be on par with the rest of the world with the help of this low-cost, reliable, and user-friendly product which can recognize text from public information boards.

References

1. Strotthe T et al (1997) Mobility of blind and elderly people interacting with computers. National Institute for the Blind, report on the MOBIC project. http://www.tiresias.org/reports/mobicf.htm
2. Real Time Text Detection and Recognition on Hand Held Objects to Assist Blind People. In: 2016 international conference on automatic control and dynamic optimization techniques (ICACDOT), International Institute of Information Technology (I^2IT), Pune
3. Venkateswarlu K, Velaga SM. Text detection on scene images using MSER
4. Islam MR, Mondal C, Azam MK, Syed A, Islam MJ Text detection and recognition using enhanced MSER detection and a novel OCR technique
5. Gómez L, Karatzas D. MSER-based real-time text detection and tracking
6. Kim KI, Jung K, Kim JH (2003) Texture-based approach for text detection in images using support vector machines and continuously adaptive mean shift algorithm. IEEE Trans Pattern Anal Mach Intelligence 25(12):1631–1639
7. Koo HI, Kim DH (2013) Scene text detection via connected component clustering and nontext filtering. IEEE Trans Image Process 22(6):2296–2305
8. Srivastav A, Kumar J (2008) Text detection in scene images using stroke width and nearest-neighbor constraints. In: TENCON IEEE region 10 conference, pp 1–5
9. Zhou G, Liu Y, Tian Z, Su Y (2011) A new hybrid method to detect text in natural scene. In: 18th IEEE international conference on image processing (ICIP), pp 2605–2608
10. Gómez L, Karatzas D (2014) MSER-based real-time text detection and tracking. In: 22nd international conference on pattern recognition (ICPR), pp 3110–3115

Compromising Cloud Security and Privacy by DoS, DDoS, and Botnet and Their Countermeasures

Martin K. Parmar and Mrugendrasinh L. Rahevar

Abstract Today, every firm either academic or private or government sectors are using the cloud as a platform to store and communicate information over the internet. Lots of data are being exchanged using various applications software and services. Services are integrated and reused the information over WWW. To effectively deal with all information, people are moving on cloud computing platform because of improving the cost of hardware and software, only pay what you use and setting up infrastructure easily and available for 24×7. Even though lots of advantages of sharing data on a cloud platform, their obvious question arise is that are those data secure and maintain its privacy by the service provider? Data privacy and security is an essential thing in today's world. And to provide security and privacy over cloud computing is often challenging part for many organizations. In this paper, we represent some security challenges for cloud platform as services. We state with two security attack using DDoS and Botnet, and also show some countermeasure.

1 Introduction

Today, use of cloud network providers is rapidly increasing in almost all the sectors of business. There is huge demand for information storage and after that, it can be shared easily anywhere anytime. Services providers of cloud are providing services as SAAS, PAAS, and IAAS [1]. Increasing usages of cloud computing also produce some issues such as security and privacy issues, availability, reliability, and so on. One of the major concerns here is the compromising cloud security and privacy by exploiting malicious attacks. In this paper, we mainly focus effect of DoS, DDoS, and Botnet attack (Fig. 1).

M. K. Parmar (✉) · M. L. Rahevar
U & P U. Patel Department of Computer Engineering, CSPIT, CHARUSAT, Changa, Gujarat, India
e-mail: martinparmar.ce@charusat.ac.in

M. L. Rahevar
e-mail: mrugendrarahevar.ce@charusat.ac.in

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_17

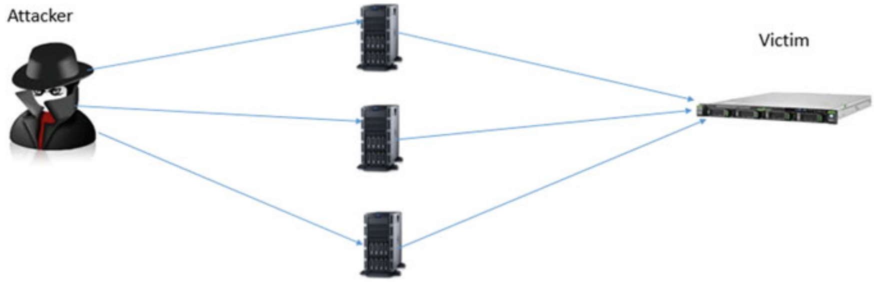


Fig. 1 DDos attack

The major effect of DDoS attacks are given below [2]:

- Unavailability of target servers.
- Increase traffic which slow down network performance.
- Significant increase in large number of emails, queries, etc.
- Make an entire network (wired or wireless) unavailable by frequent disconnection.

1.1 Advantages of Cloud Computing

- **Flexible costs**: Compare to traditional computing where a company needs to purchase expensive hardware and software whether it is required or not in future, by setting up cloud architecture, one can demand hardware and software as per requirements and also easily upgrade the system. So by avoiding large expenditure on hardware and software, a cloud can improve cost by matching your current requirements.
- **On-demand**: Looking toward current requirements as well as future requirements, your cloud environment grows. Adding and removing cloud service is on-demand [3].
- **Focus on your business**: It is difficult to monitoring infrastructure 24/7 which is time-consuming and expensive. A cloud can manage solution like monitoring your infrastructure and keeping your data secure. In addition, a cloud can provide creative and practical solutions to keep your IT infrastructure working effectively.
- **Availability of Data**: Information on a cloud is available every time anywhere. A user can access and operate data which is available in cloud storage safely using a mobile phone or computer.

1.2 Disadvantages of Cloud Computing

Technical outages

Cloud services entirely depend on Internet connectivity. You can get services efficiently if your Internet connectivity is proper. You would not be able to access any of your applications, resources or information from the cloud, if your Internet connection is offline.

Cloud Security and Privacy

By using cloud computing environment, we give our valuable information to cloud service providers. All the confidential information of the company is now stored in cloud storage and also can be accessed from anywhere using the internet. Thus, trusting on service providers are the biggest question. To minimize the risk of trusting on third-party cloud service providers, we should choose appropriate and the most reliable cloud service provides.

Interoperability

Even if cloud vendor gives surety about reliable and flexible cloud environment, it has been found so difficult to migrate or change service providers. As the infrastructure and storage structure is different than other service providers.

Security Issues in Cloud Computing

Data security is a big concern. There are lots of data stored on a cloud platform. Moreover, these data are shared or exchanged among different consumers over the heterogeneous environment. That is why the security, as well as privacy of user's data, should be maintained in a proper way. Even though there are enough service level agreements (SLAs) between provider and consumer but still there are certain issues such as authentication, authorization, confidentiality, loss of information, security breaches, and other vulnerability in the system [4]. These issues may be caused by internal or external intruders who willingly or not but causes serious damages and result in the loss of a piece of information.

There are mainly two issues exist with cloud security and privacy.

1. No any control over data
2. Dependency on the cloud service provider.

Issues due to no control over data are given below:

1. There is lack of transparency between user and service provider. Because users may not be aware of their data which are going to use by providers and how, when, why, and where they will use [5].
2. Sensitive data such as sharing private information over social media applications like videos or photos. Most providers often do mining or research of user data. In this case, also users may not aware of sharing of their private information among customer of providers.
3. There are certain applications in mobile devices which are driven by cloud providers and without noticing of users, their private information may be used or shared. So, the users are always in a risk about the transparency of their data since they think all information is stored locally between mobile and servers.
4. Sometimes, the connection between the cloud provider and consumer are compromised by intruders by eavesdropping sensitive information. There are also

DNS spoofing and denial of service attack which often make end resources available for users.
5. Entire control is shifted to cloud providers so trust management between cloud providers and customers are very hard to believe. If a customer wants to make some verification, then appropriate login should be provided.
6. There are so many data scattered around and one important aspect is that delectation of those data. For example, picture or image type of data, there should be some protection mechanism or guarantee that once data will be deleted and then it will not exist any other place.
7. Cloud computing should be provided through secure and reliable communication to end user. There should not be any other third-party involve between provider and service consumers.

Issues with the dependency on the Cloud Service provider are given below:

1. The availability is the biggest issue. If the services which are provided by cloud computing providers are down, then it would be difficult for customers for accessing data. Shutting down the services create so many problems and make data inaccessible.
2. There is not any contract between users and some service provides, therefore, users are unaware of some misuse of information [5].
3. Cloud services are provided by so many services providers which target different customers sometimes, so it is difficult to change providers.

2 Attacks on Cloud Computing

2.1 DoS Attack

Denial of service (DoS) is a kind of attack in which the end resources will become inaccessible and unavailable to end users. Unlike, other attacks in which your secure data is being stolen or eavesdropping and intercepting any information, DoS is only intended to make your end resources such as server or website or any other resources, unavailable [6].

DoS Attack Prevention:

There are some steps to follow for prevention of DoS attack:

- Monitor your HTTP traffic and find unexpected behavior by installing a firewall.
- Keep an eye on social networking or any other discussion forum which gives hints about incoming DoS attack.
- Use third-party service which helps to minimize DoS attacks.
- Make your firewall policy strong to block unwanted traffic.

2.2 DDoS Attack

Distributed denial of service (DDoS) attack tries to make an end resources or a network resource unattainable to users. A DDoS attack uses multiple connected devices which are often called botnets. Using botnets, they interrupt the end resources and make them down [7].

There are basically two types of DDoS attack

1. Application layer DDoS attack
   Main aim of the attacks is sending n number of requests to target machine which makes the end machine overloaded. Furthermore, it causes high processing power of CPU and memory usage that conclusively suspend or fail down the application. Application layer attacks like HTTP floods, slow attacks, and zero-day assaults which targets vulnerabilities in operating system, websites and communication protocols.
2. Network layer DDoS attack
   Using botnets, these attacks are typically launched. The main goal behind network layer attack is to expend the end resource's upstream bandwidth and result in network saturation.
   Network layer attacks like UDP floods, SYN floods, DNS amplification and IP fragmentation are typically used to overwhelm server by sending number of packages.

DDoS attacks can be performed by making target infrastructures and services such as DNS servers unavailable to users. It makes false DNS request by botnet devices.

DDoS Attack Prevention:

On-Premise Appliances

The on-premise approach to DDoS protection uses hardware appliances deployed inside a network, placed in front of protected servers. Such appliances usually have advanced traffic filtering capabilities. Typical mitigation appliances can be effectively used to filter out malicious incoming traffic. This makes them a viable option for stopping application layer attacks.

However, several factors make it unfeasible to rely on appliances:

Scalability remains an issue. The ability of the hardware to handle large amounts of DDoS traffic is capped by a network's uplink, which is rarely more than 10 Gbps (burst).

On-premise appliances need to be manually deployed to stop an attack. This impacts time to response and mitigation, often causing organizations to suffer downtime before a security perimeter can be established.

Finally, the cost to purchase, install and maintain hardware is relatively high—especially when compared to a less costly and more effective cloud-based option. This makes mitigation appliances an impractical purchase unless an organization is obligated to use on-premise solutions (e.g., by industry-specific regulations).

Off-premise, Cloud-based Solutions

Off-premise solutions are either ISP-provided or cloud-based. ISPs typically offer only network layer protection, while cloud-based solutions provide additional filtering capabilities required to stop application layer attacks. Both offer virtually limitless scalability, as they are deployed outside of a network and are not constrained by the previously identified uplink limitations.

Generally, off-premise mitigation solutions are managed services. They do not require any of the investment in security personnel or upkeep required by DIY solutions and on-premise hardware. They are also significantly more cost-effective than on-premise solutions while providing better protection against both network and application layer threats.

Off-premise solutions are deployed either as an on-demand or always-on service, with most market-leading vendors offering both options.

On-demand option

Enabled by BGP rerouting, the on-demand option stops network layer attacks—including those directly targeting the origin server and other components of core network infrastructure. These include SYN or UDP floods, which are volumetric attacks designed to clog network pipes with fake data packets.

Always-on option

The always-on option is enabled through DNS redirection. It stops application layer assaults attempting to establish TCP connections with an application in an effort to exhaust server resources. These include HTTP floods, DNS floods and various low and slow attacks (e.g., Slowloris).

2.3 Comparison of DoS Versus DDoS [8]

Following are the comparison about effects of DoS and DDoS attack.

DoS	DDoS
✓ Use dedicated computer or Internet connection to attack	✓ Use multiple machines and connections to attack
✓ Target system is overloaded by sending many request by one machine	✓ Target systems are overloaded by sending many requests by different machines
✓ Compare to DDoS, it is less difficult as attack is launched from one machine	✓ It is difficult to identify legitimate traffic than attacker's traffic as attacks are launched from different machines
✓ There is no any malware there	✓ Using BotNets, thousands of machined can be infected

Finally, DoS attack can be stopped using right security policy but DDoS attack can become a headache to identify legitimate traffic.

2.4 Botnet DDoS Attacks

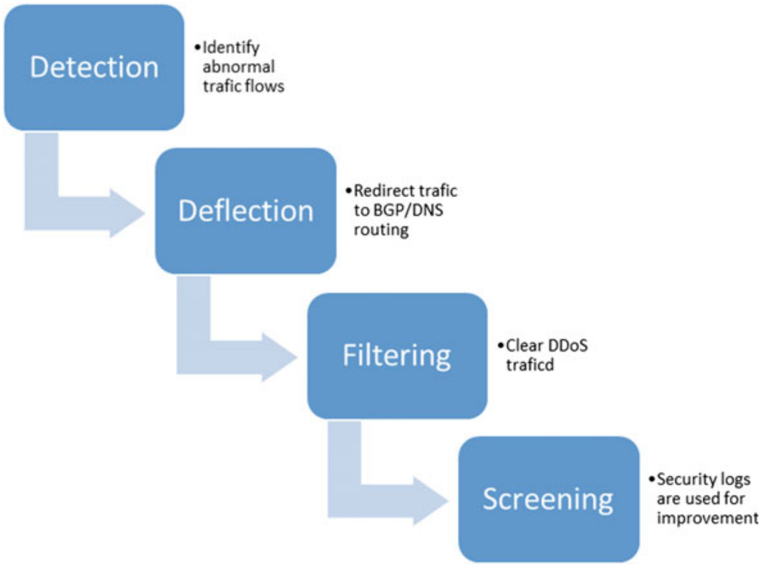
It is also known as "Zombie Army". A botnet attack is launched from the group of connected machines with the internet. It mainly infects machines by injecting malicious malware in distributed or heterogeneous environment. Once all the machines become infected, it is called bot and then, the entire network can be controlled by remote location [9].

A botnet attack can be performed by different purpose such as information stealing (financial, user credentials for login, and relive any online identity). The botnet attack is mostly used with DDoS attack in which it tries to infect as many machines as possible and break down the entire network or target machines without knowledge of user [10].

3 Proposed Model for Mitigating DDoS Attack

The DDoS mitigation basically protects a destination target from a distributed denial of service (DDoS) attack.

A DDoS mitigation process can be broadly defined by following four stages:



1. **Detection**: In this stage, the main focus is on recognizing different stream of traffic diversity which triggers the action for DDoD attack. It is up to an individual who may have the skill to identify an attack impressively. It would be great if the attack is detected instantly.
2. **Deflection**: After detection, the traffic is filtered and discarded or it is diverted from target machine.
3. **Filtering**: By observing and analyzing, one can recognize particular patterns which will make differentiate among legitimate traffic, DDoS-Bot and malicious request. The main thing is that to provide complete transference by identifying, detecting and block attack without any knowledge of users.
4. **Screening**: It is a very crucial point to review all possible security logs and collect detail information for any malicious action. It is good practice which can improve to identify further attack in future.

3.1 Network Layer Mitigation Techniques

Various service providers have different methods of protecting from network layer DDoS attacks, some of which are not preferable than others:

- **Null routing**: In normal routing, each package directs to destination IP but in null routing, data goes nowhere. It means null routing contains IP address which does not exist. This way, it creates lots of junk data in network.
- **Sinkholing**: Sinkhole redirects traffic from real destination machine to another machine which is known as sinkhole. It is used for good purpose and also bad. Sinkhole measures data flow in the network so it can be used as a research tool. But, sometimes, attacker can redirect traffic from legitimate user.
- **Scrubbing:** An improvement on an arbitrary sinkhole, scrubbing routes all ingress traffic through a security service. Malicious network packets are identified based on their header content, size, type, a point of origin, etc. The challenge is to perform scrubbing at an inline rate without causing lag or otherwise impacting legitimate users [11].

3.2 Application Layer Mitigation Techniques

Being much stealthier than their network layer counterparts, application layer DDoS attacks typically mimic legitimate user traffic to evade security measures. To stop them, your solution should have the ability to profile incoming HTTP/S traffic, distinguishing between DDoS bots and legitimate visitors.

During mitigation service trials, testing its application layer defenses is essential. Effective filtering uses cross-inspection of HTTP/S header content and behavioral patterns, in addition to IP and autonomous system number (ASN) information. Many security services also use different types of challenges, such as testing each request for its ability to parse JavaScript and hold cookies [12].

It is equally important to verify that the service does not overuse CAPTCHAs, “delay pages” and other such filtering methods that only serve to annoy legitimate visitors.

3.3 Mitigating Botnet DDoS Attacks

3.3.1 Network Layer Attacks

In the event of a network layer attack, we can prepare a model which provides dynamic resource overprovisioning—offering nearly limitless, on-call scalability. In addition to, using reverse proxy helps mask original service IPs which is the first line of protection against direct-to-IP attacks [13].

In an event where target IPs are already known, and the masking effect is insufficient,

1. Enforce routing policies using BGP announcements which ensure that all incoming traffic travels through legitimate server first, where it undergoes deep packet inspection and filtration process [14].
2. After first step, only clean traffic is permitted to reach the source via a secure, two-way GRE tunnel.
3. During the mitigation process, the security system continually monitors the IP addresses and sortie patterns.
4. Lastly, suspicious data is transmitted to DDoS threat database and this will benefit to all of service consumers. Because of that, it generates quick response for any malicious activity and also prepare to identify for any future botnet attack [15].

3.3.2 Application Layer Attacks

Analyzing and observing traffic data from legitimate, behavior of network, and number of request to the server can be useful for mitigating application layer DDoS attack.

3.3.3 DNS-Targeted Attacks

Webmasters can set their authoritative domain name server, while DNS zone file management remains independent of the network.

- All inbound DNS queries first reach to legitimate server, where malicious requests are automatically filtered out. Only legitimate ones are allowed to pass through, enabling smooth traffic flow at all times.

4 Conclusion

Cloud computing has tremendous advantages of data storing and make data available every time. Service providers are providing effective solution of on-demand services at reduced cost. Thus, every enterprise, private sectors, government sectors, and other firms use cloud platform for their business growth. As a result, there is obvious question might be raised about certain issues like information security and privacy. Moreover, attackers always try to do malicious activities such as make end system unavailable by overloading server or infect the thousands of machines by injecting malwares. In this paper, we have demonstrated some issues of unavailability of resources using DoS, DDoS, and Botnet attacks. Dos attack can be launched from one machine and by applying proper security policy, it can be prevented. But, attacking from multiple sources particularly in cloud environment where every resource is geographically distributed and in addition, identifying legitimate resources from bots, are very difficult and time-consuming task.

In this paper, we presented a proposed model for mitigating DDoS attack and DDoS using Botnet attack. By using detection, deflection, filtering and screening, we can lower down the change of DDoS attack and also caught fake requests from bots. But still, it is difficult to deal with number of fake requests using bots and identifying right controller who command the bots. So in the future, we can use such an effective technique to handle the number of bots in cloud environment.

References

1. Deshmukha RV, Devadkar KK Understanding DDoS attack & its effect in cloud environment. In: 4th international conference on advances in computing, communication and control (ICAC3'15)
2. Rai A, Challa RK (2016) Survey on recent DDoS mitigation techniques and comparative analysis. In: Second international conference on computational intelligence & communication technology. IEEE, New York
3. Wang B, Zheng Y, Lou W, Hou YT (2015) DDoS attack protection in the era of cloud computing and software-defined networking. www.elsevier.com/locate/comnet (Elsevier)
4. Somani G, Gaur MS, Sanghi D, Conti M, Buyya R (2017) DDoS attacks in cloud computing: issues, taxonomy, and future directions. www.elsevier.com/locate/comcom (Elsevier)
5. Zissis D, Lekkas D (2012) Addressing cloud computing security issues. Future Gen Comput Syst 28(3):583–592. https://doi.org/10.1016/j.future.2010.12.006
6. Balobaid A, Alawad W, Aljasim H (2016) A study on the impacts of DoS and DDoS attacks on cloud and mitigation techniques. In: 2016 international conference on computing, analytics and security trends (CAST) College of Engineering Pune, India. Dec 19–21, 2016
7. Alosaimi W, Zak M, Al-Begain K (2015) Denial of service attacks mitigation in the cloud. In: 9th international conference on next generation mobile applications, services and technologies
8. Incapsula. https://www.incapsula.com/ddos/denial-of-service.html
9. Yan Q, Yu FR (2015) Distributed denial of service attacks in software-defined networking with cloud computing. In: Security and privacy in emerging networks. IEEE, New York
10. Liu X, Yang X, Lu Y (2008) To filter or to authorize: network-layer DoS defense against multimillion-node botnets. In: ACM SIGCOMM computer communication review. ACM, New York

11. Bhardwaj A, Subrahmanyam GVB, Avasthi V, Sastry H, Goundar S (2016) DDoS attacks, new DDoS taxonomy and mitigation solutions—a survey. In: International conference on signal processing, communication, power and embedded system (SCOPES). IEEE, New York
12. Khor SH, Nakao A (2011) DaaS: DDoS mitigation-as-a-service. In: IPSJ international symposium on applications and the internet. IEEE, New York
13. Bawany NZ, Shamsi JA, Salah K (2017) DDoS attack detection and mitigation using SDN: methods, practices, and solutions. In: Review article—computer engineering and computer science. Springer, Berlin
14. Wong F, Tan CX (2014) A survey of trends in massive DDoS attacks and cloud-based mitigations. Int J Netw Secur Appl (IJNSA) 6(3):57–71
15. Farahmandian S et al (2013) A survey on methods to defend against DDoS attack in cloud computing. Proc Recent Adv Knowl Eng Syst Sci 6:185–190

A New Automated Medicine Prescription System for Plant Diseases

S. Sachin, K. Sudarshana, R. Roopalakshmi, Suraksha, C. N. Nayana and D. S. Deeksha

Abstract In the current situation, agriculture is facing a wide number of problems to address the increasing global population. Also, the plant diseases affect the production and quality of crops. Specifically, plant disease severity identification is the most important problem in the agricultural field which can avoid the excess use of pesticides and minimize the yield loss. In the existing systems, no methodology exists to identify the disease severity and to prescribe the required quantity of medicines to be sprayed. In order to solve this problem, an automated medicine prescription system is proposed in this paper, which takes the images from the uncontrolled environment, enhances, and preprocesses the images received for the identification of disease. Precisely, in the proposed framework, k-means and SVM algorithms are used for clustering and disease identification tasks, respectively. Experimental setup and snapshots of results demonstrate the performance of the proposed system, by means of indicating the severity of the identified disease.

1 Introduction

India is an agricultural country, wherein a large portion of the population relies upon farming. Indian economy mainly stands on agriculture, since over 58% of income comes through agricultural segment [1]. In the current situation, agriculture faces wide number of problems due to the increasing global population and the plant diseases affecting the production and quality of the crops. However, agriculture is influenced by various other climatic factors such as drought, inordinate rainfall, and

S. Sachin · K. Sudarshana (✉) · R. Roopalakshmi · Suraksha · C. N. Nayana · D. S. Deeksha
Alvas Institute of Engineering and Technology, Moodbidri, Mangaluru 574225, India
e-mail: kerenalli@gmail.com

S. Sachin
e-mail: shettysachin1996@gmail.com

R. Roopalakshmi
e-mail: drroopalakshmir@gmail.com

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_18

sudden changes in temperature, increase in pollution factors, and release of industrial as well as chemical wastage into agricultural lands [2].

Out of many factors which affect the production, the plant diseases are the more significant. Hence, detecting the plant disease plays a more significant role. Plant diseases are the main reason for the reduction in the agricultural product quality [3]. Disease is a sick situation of the healthy state of the plant by which major biological process of the plant life. Current methods for detecting the plant diseases require the naked eye observations by the experts, which is inefficient, costlier, and time-consuming.

Farmers in many countries do not have adequate facilities or even ideas that they can contact to experts, which makes the consulting experts even cost expensive. In the current context, there a need for monitoring the large forms automatically. Machine vision-based detection of the diseases symptoms on the plant leaves makes the disease severity detection easier and cost-effective [4]. In this way, it is possible to automatically detect plant diseases using plant leafs images for various image processing operations.

The paper is further organized as follows: Sect. 2 discusses about existing work, and in Sect. 3, a description on an integrated novel technique for automated medicine prescription system is proposed and implemented. Section 4 concludes the current work followed by future enhancements.

2 Related Work

Most of the initial symptoms of plant diseases are microscopic and hence diagnosis is limited by the human vision. Recently, in 2015, Pujari et al. [5] presented a study on the image processing techniques. In this work, symptoms of various fungal diseases are affected on different crops. The above method is complex and suffers from various in terms of high variability in outdoor conditions and general symptoms.

Khirade and Patil [6] used image processing techniques for diseased part of the plant leaf segmentation, where the features of infected leaf are extracted and classified. However, the accuracy was less. In 2012, Revathi and Hemalatha proposed a HPCCDD using leaf spot images. They are captured by farmers through smartphones [7]. The Sobel and Canny filters were used to identify the edges. The HPCCDD algorithm is used for image analysis and disease classification. It used only single leaf images to obtain the disease severity.

In 2013, Husin et al. [8] described a technique in which chilly plant diseases visual symptoms are identified using the analysis of colored images. The input image was enhanced to preserve information of the affected pixels before extraction. This method takes more time to detect the disease and hence failed to render the expected performance.

Sanjay and Nitin [9] proposal was based on k-means clustering to solve low-level image segmentation. SGDM method was used for extracting statistical texture features. İsolating the picture from the unwanted scenes was difficult. In 2009, Smith and Camargo [10], set out an image processing method to identify the visual symptoms of plant diseases from colored images analysis and it fails to extract the diseased region parameters and image classification.

In 2012, Di Cui et al. [11], proposed method for multispectral images. It is used to identify the plant leaf rust, along with disease severity. The parameters used were color distribution and ratio of infected area (RIA) or rust color index (RCI). A reference database can be built for automated rust severity detection.

Similarly, Tian and Zhang [12] used hyperspectral imaging for comparative examination on downy mildew disease of cucumber plant.

Without bothering the illness level, the pesticides are sprayed unconditionally. In 2017, Mude et al. [13] described a technique for the pesticide spraying only of plant got affected; otherwise not. Here, picture handling procedures were applied to discover the illness. It also suffers from less severity identification.

Automatic spraying of pesticides shall reduce the health issues caused by the overexposure to farmers. In 2012, Husin et al. [14] proposed a method to detect cotton leaf diseases efficiently using image processing techniques. A tool implemented in MATLAB is used to detect types of diseases on leaf. A robot is used to spray the pesticides once the disease is detected.

To outline, the existing literature fails to focus on disease severity, accuracy of results, and time complexity. In order to overcome from the above issues, we propose a disease detection model.

3 Proposed Framework

Figure 1 represents the proposed system, which consists of six phases, namely image acquisition, image enhancement, preprocessing, segmentation, feature extraction, classification, and the seventh step is medicine spraying. In the proposed framework, image is acquired through the camera which is associated with the robot. The robot movement is controlled by the Android application through Bluetooth. On the other hand, it also controls the camera movement to capture the images. The view that is created by the camera is shown on the system to catch the proper parts of the plant.

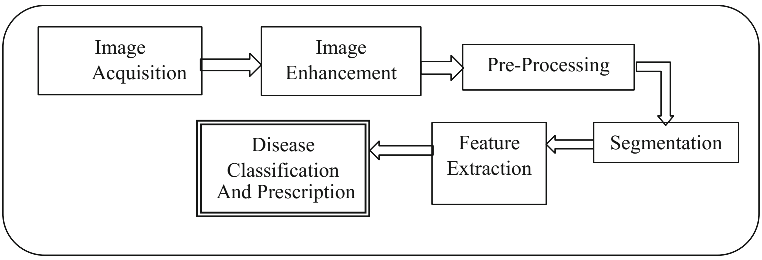


Fig. 1 Block diagram of proposed automated medicine prescription system

In the preprocessing step, the input RGB image is changed over into grayscale image. The contrast of the image is improved using adaptive histogram equalization. After the enhancement, noise is removed from the input image. The resultant images are utilized for further processing. K-means clustering algorithm is used to form different clusters. In this method the images are divided into three clusters in which cluster containing affected region is considered for further processing. Gray-level Co-occurrence Matrix is utilized as a part of the proposed model for the feature extraction, which extracts 13 features including mean, standard deviation, entropy, RMS, variance, smoothness, kurtosis, skewness, IDM, contrast, correlation, energy, and homogeneity, which are used for the classification process. SVM algorithm is used in this method to classify the disease. It maps the extracted features with the database stored to identify the disease type. Once the disease type is identified, its severity can be identified by using the following formula,

$$\text{Disease severity} = \text{Affected pixels/Total pixels} \tag{1}$$

Once the disease severity is identified, then it is compared with the database information to identify the pesticide and quantity of pesticides to be sprayed over the affected region of the plant. The same information shall be forwarded to the robot, which will find the required pesticide and spray it over the affected region.

4 Experimental Setup

Figure 2 depicts the experimental setup of different components of the system and their respective connections. Arduino Uno-based control system handles the activities of the robot and pesticide spraying. Arduino is connected to servo motor to control the camera.

Arduino receives the instruction from the Android application through Bluetooth. L293D motor driver is connected to the Arduino to handle the robot movement. Relay is connected to various pesticides which handles the spraying system.

5 Results and Discussion

Figure 3 represents the snapshot of preprocessing of diseased leaf images. The captured leaf image is loaded and its contrast is enhanced. In visual perception, contrast is determined by the difference in the color and brightness of the object with other objects. After this step image segmentation takes place, in which the leaf image is divided into multiple clusters and the cluster that is needed for the processing is selected. Figure 4 shows the snapshot of clustering of the leaf images. Clustering is used to separate the group of objects. It also helps to separate the healthy region of the image from the diseased one. One resultant image is selected to identify the type

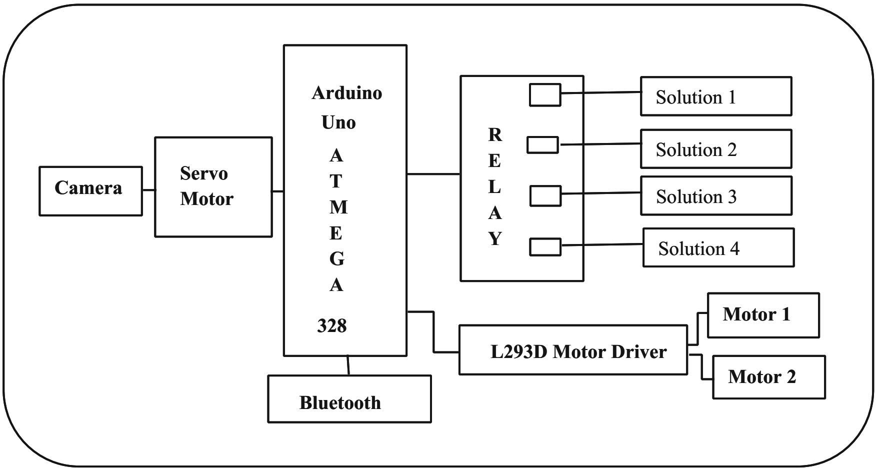


Fig. 2 Experimental setup

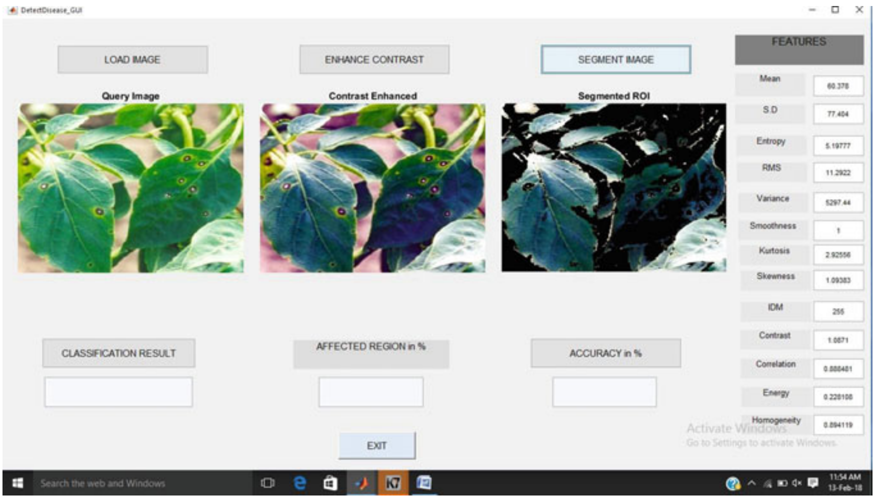


Fig. 3 Snapshot showing preprocessing of diseased leaf

of disease. In Fig. 4, the original image which is divided into three clusters and the cluster that contains the diseased part is selected for the further processing. Figure 5 shows the snapshot of disease severity details of the diseased leaf image. Precisely, features that are necessary for the identification of the disease are extracted from the segmented image. These features are shown in Fig. 5.

Once the features are extracted, disease is identified in Fig. 5 and it identifies the disease which is popularly known as Cercospora leaf spot and its severity is

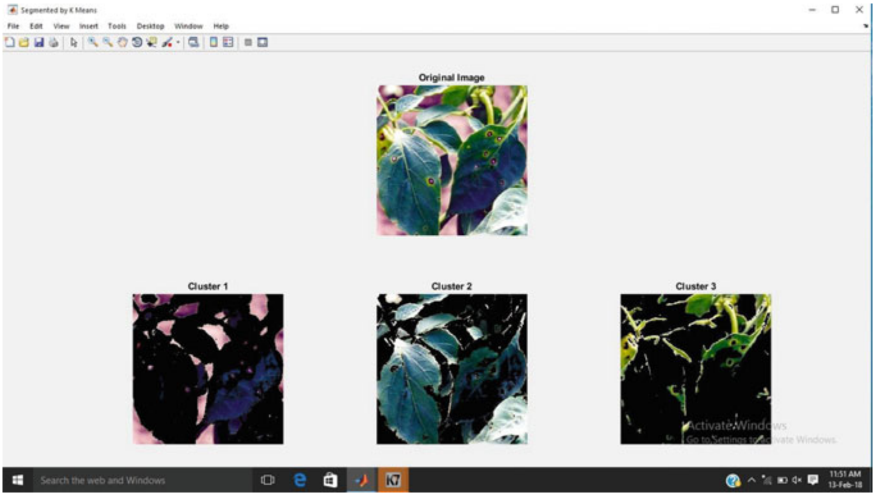


Fig. 4 Snapshot showing clustering of the diseased leaf image

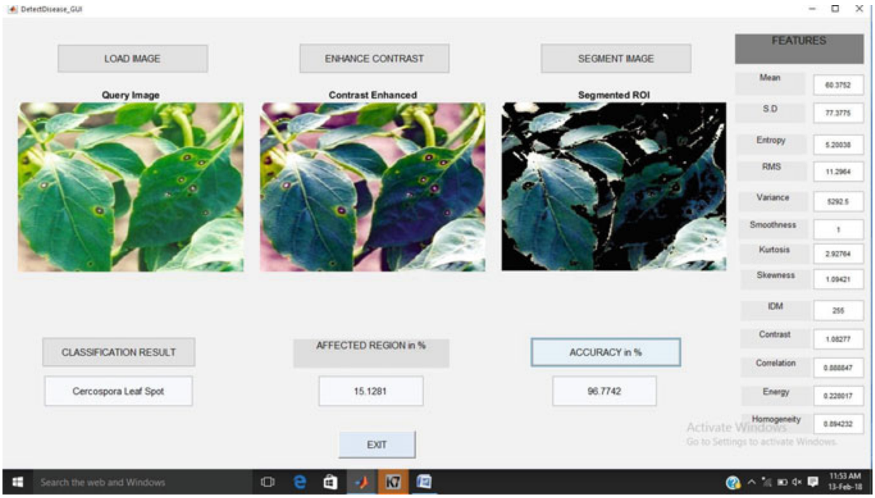


Fig. 5 Snapshot showing disease severity details of the diseased leaf image

15.1281%. Accuracy of the identified leaf disease is 96.77%, which illustrates the good performance of the proposed framework.

6 Conclusion

The proposed framework is useful for farmers and horticulturists by providing useful information related to plant leaf diseases and their identification techniques. The proposed system, can also control, schedule, and monitor all the agricultural recourses, which offers a great aid to reduce the time and cost to the farmer. This system helps the farmers, by reducing their efforts and also increases the production of crops. This method can be further enhanced, by incorporating the features such as multiple plant multi-disease detection system and developing a system for long-distance disease detection.

References

1. Agriculture in India: Information about Indian agriculture its importance. http://www.ibef.org/industry
2. Factors that affect the distribution of agriculture. https://www.s-cool.co.uk/a-level/geography/agriculture/revise-it/factors-that-affect-the-distribution-of-agriculture
3. The impact of plant disease on food security. www.mdpi.com/journal/agriculture/special_issues/plant_disease
4. Singh V, Misra AK (2017) Detection of plant leaf diseases using image segmentation and soft computing techniques. Elsevier Inf Process Agric 4(1):41–49
5. Pujari JD, Yakkundimath R, Byadgi AS (2015) Image processing based detection of fungal diseases in plant. In: Elsevier international conference and communication technologies (ICICT), pp 1802–1808
6. Khirade SD, Patil AB (2015) Plant disease detection using image processing. In: Proceedings of international conference on computing communication control and automation, pp 768–771
7. Revathi P, Hemalatha M (2012) Classification of cotton leaf spot diseases using image processing edge detection techniques. In: Proceedings of IEEE international conference on emerging trends in science, engineering and technology, pp 169–173
8. Husin Z, Md Shakaff AYB, Aziz AHBA (2013) Feasibility study on plant chilli disease detection using image processing techniques. In: Proceedings of international conference intelligent systems, modelling and simulation. IEEE, New York, pp 291–296
9. Sanjay BD, Nitin PK (2013) Agricultural plant leaf disease detection using image processing. Int J Adv Res Electr Electron Instrum Eng 2(1):599–602
10. Camargo A, Smith JS (2009) An image-processing based algorithm to automatically identify plant disease visual symptoms. Elsevier Biosyst Eng 17(1):9–21
11. Cui D, Zhang Q, Li M, Hartman GL, Zhao Y (2010) Image processing methods for quantitatively detecting soybean rust from multispectral images. Elsevier Biosyst Eng 22(4):186–193
12. Tian Y, Zhang L (2012) Study on the methods of detecting cucumber downy mildew using hyperspectral imaging technology. Elsevier Int Conf Phys Biomed Eng 11(5):743–750
13. Mude S, Naik D, Patil A (2017) Leaf disease detection using image processing for pesticide spraying. Int J Adv Eng Res Dev 4(4):1–5
14. Husin ZB, Md Shakaff AYB, Aziz AHBA, Farook SM (2012) Plant chilli disease detection using the RGB color model. In: Proceedings of research notes in information science (RNIS), pp 291–296

A Comparative Assessment of Segmentations on Skin Lesion Through Various Entropy and Six Sigma Thresholds

Srinivasan Sankaran, Jason R. Hagerty, Muthukumaran Malarvel, Gopalakrishnan Sethumadhavan and William V. Stoecker

Abstract We present four entropy-based methods for colour segmentation within a lesion in a dermoscopy image for classification of the image as melanoma or benign. Four entropy segmentation methods are based on Tsallis, Havrda and Charvat, Renyi and Kapur entropy measures. Segmentation through Six Sigma threshold as preprocessor is also evaluated by this assessment approach. The proposed methods are inspired by two clinical observations about melanoma. First, colours within a lesion provide the most useful measures for melanoma detection; second, the disorder in colour variety and arrangement provides the best assessment of melanoma colours. These observations lead to the hypothesis that colour disorder is best measured by entropy. The five different models for colour splitting studied with SSIM measures taken from each region in the colour-split image for segmentation assessment. Based on the score helps to understand segmentation region assessment effectively.

1 Introduction

Melanoma causes most skin cancer deaths, with an estimated 9730 people expected to die of melanoma in the United States in the year 2017 [1]. Melanoma is fully curable if diagnosed early. The number of benign lesions biopsied exceeds melanomas by over a factor of 10, yet many cases of melanoma are incorrectly diagnosed by dermatologists [2–4].

The dermoscopy imaging method has been reported to be an important tool in the early detection of melanoma [5–8]. Studies have shown that dermoscopy increases the diagnostic accuracy over clinical visual inspection in the hands of experienced

S. Sankaran · M. Malarvel · G. Sethumadhavan (✉)
SASTRA Deemed University, Tirumalaisamudram, India
e-mail: sgk@mca.sastra.edu

J. R. Hagerty · W. V. Stoecker
S&A Technologies, Rolla, MO, USA

S. Sankaran
HCL Technologies Limited, Chennai, India

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_19

physicians [9–11]. Even higher diagnostic accuracy has been reported recently for automated digital image analysis [3, 4].

Colours are believed to provide the most critical features in melanoma diagnosis, especially when combined with a location in the lesion to capture colour chaos [12–14]. Analytic colour descriptors usually employ the red, green and blue (RGB) colour space. Of all analytic descriptors for melanoma region detection, these colours were evidenced as the most significant descriptors. Ferris et al. evidenced that the top three features through statistical derivatives for melanoma were all colour descriptors from colour histograms and colour asymmetry [4]. Rubegni et al. found that the top critical feature for melanoma detection was red asymmetry [15], capturing the “colour island” chaos [16]. Other colour features for melanoma detection are a variation of hues [17], analytical colour analysis of variegation [18], RGB colour channel statistical parameters [19], spherical and La*b* colour coordinate features [20], and the number of colours of concern present within the skin lesion [21]. Colour quantization was performed using the median split colour algorithm [22]. Crisp and fuzzy colour histogram techniques were used to identify melanoma colours [23, 24]. The improved colour contrast was used to better separate colours [25]. A colour palette mimicking that used by dermatologists was developed [26].

Based on Otsu and various entropy functions, image segmentation at each possible threshold in the grey image is investigated through randomness measure for its evaluation [27]. The idea of multilevel segmentation was a key item conceived through this paper. Based on the analysis, results show an outperforming behavior of generalized entropy measure of Havrda function in comparison with any other existing entropy functions. Sankaran et al. [28] developed segmentation using Six Sigma for malarial parasitaemia image and derived infectious cells based on Kapur entropy measure.

In this paper, four entropy-based colour splitting and one statistical method to capture colour variety and arrangement are presented. The rest of the paper is organized as follows. Section 2 explains the entropy-based colour segmentation algorithms used in this research. Section 3 discusses segmentation generated using Six Sigma as a threshold. Section 4 explains the assessment methodology. Section 5 elaborates about tool and methodology applied. Section 6 describes the results and discussion.

2 Entropy-Based Threshold Segmentations

A total of four different entropy-based colour segmentation algorithms are used in this study. Entropy is a measure of uncertainty associated with information of a source. Secretion of melanin in the affected skin region is being measured using the amount of colour variation present in that region; and this colour distortion is also highly uncertain due to the inherent property of the disease. In general, distribution of uncertainty data can be assessed through probability distribution using the normalized histogram on various colour channels. Based on the colour distortion of colour channels in melanoma images, various entropy methods guide to derive appropriate threshold to segment the lesion region in the image. This uncertainty in colour spread information on various channels is sourced for Tsallis entropy [29], which is defined as

$$S_{\propto} = \frac{1}{\propto - 1}\left(1 - \sum_{i=1}^{n} p_i^{\propto}\right) \quad (1)$$

where P_i is the normalized histogram defined as $P_i = n_i/n$ and n be the total number of pixels and n_i be the number of pixels with grey value.

Let $f(x,y)$ be the image of spatial resolution $M \times N$ and bit level resolution k with the grey scale [0, $L - 1$] where $L=2k$. Let, $t = 0$ to $L - 1$ be the probability distribution with the pixel values ranging from 0 to t as object pixels (O) and the pixels values ranging from $t+1$ to $L - 1$ as background pixels (B). Then,

$$H(O, t) = -\sum_{i=0}^{t} P_i \ln(P_i) \quad (2)$$

$$H(B, t) = -\sum_{i=t+1}^{L-1} P_i \ln(P_i) \quad (3)$$

Image histogram carries important information about the content of an image and can be used for discriminating the abnormal tissue from the local unhealthy background. Havrda and Charvat have defined entropy [30] of a discrete finite probability distribution as follows:

$$H = \frac{1}{1-\alpha}\left(\sum_{i=0}^{L-1} P_i^{\alpha} - 1\right) \quad (4)$$

where α is a positive real parameter used to derive suitable threshold using this nonlinear equation. Thus, a one-parameter generalization of Shannon's entropy is Havrda and Charvat entropy. Also, it can be shown that Shannon's entropy is the limiting case of Renyi's entropy, when $\alpha \rightarrow 1$.

$$H(O, t) = \sum_{i=0}^{t} \frac{1}{1-\alpha}\left(\sum_{i=0}^{t} P_i^{\alpha} - 1\right) \quad (5)$$

$$H(B, t) = \sum_{i=t+1}^{L-1} \frac{1}{1-\alpha}\left(\sum_{i=t+1}^{L-1} P_i^{\alpha} - 1\right) \quad (6)$$

If *foreground* and *background* pixels have the same grey level histogram, then the parameter α is critical in finding Havrda and Charvat and Renyi entropy thresholds.

The typical melanoma image has colour variation in both the lesion and background skin. The Kapur threshold [31] divides the image histogram into two probability distributions, one for lesion the other for the background. The Kapur entropy threshold maximizes the sum of entropies of background and lesion.

$$H(O,t) = -\sum_{i=0}^{t} \frac{p_i}{P_t} \ln \frac{p_i}{P_t} \tag{7}$$

$$H(B,t) = -\sum_{i=t+1}^{L-1} \frac{p_i}{1-P_t} \ln \frac{p_i}{1-P_t}; \quad \text{where } P_t = \sum_{i=0}^{t} p_t \tag{8}$$

The Rényi [32] entropy is important in statistics as it provides an index of diversity, which can measure the greater number of colours found in melanomas [14, 26]. Renyi entropy sums the probability of pixel grey levels for foreground and background. The probabilities of the pixels at each grey level are combined, as in Eq. (9) using mathematical operations. The final threshold is the maximum priori combination value.

$$R = \frac{1}{1-\alpha} \ln\left(\sum_{i=0}^{L-1} P_i^{\alpha}\right) \tag{9}$$

where $\alpha(\neq 1)$ is a positive real parameter which proportionately determines the suitable threshold value.

$$H(O,t) = \sum_{i=0}^{t} \frac{1}{1-\alpha} \ln\left(\sum_{i=0}^{t} P_i^{\alpha}\right) \tag{10}$$

$$H(B,t) = \sum_{i=t+1}^{L-1} \frac{1}{1-\alpha} \ln\left(\sum_{i=t+1}^{L-1} P_i^{\alpha}\right) \tag{11}$$

$E(t)$ can be derived as the sum of $H(O,t)$ and $H(B,t)$. And the corresponding maximum value of $E(t)$ received as l_E^* where $E = T_S$, *HC, K and R* are the selected Tsallis, Havrda and Charvat, Kapur and Renyi entropy-based threshold, respectively. Equations 12 and 13 states the procedure in getting these suitable thresholds.

$$E(t)(Ts(t) = HC(t) = K(t) = R(t) = H(O,t) + H(B,t) \tag{12}$$

$$l_E^* = \max_{0 \le t \le L-1} E(t) \tag{13}$$

3 Segmentation Through Six Sigma Threshold

Dr. W. A. Shewhart proposed that variability does exist on all repetitive process through Control Chart. This methodology is applied to lesion region where variability does exist. For this purpose, to generate samples, the image is split into a rectangular window of sizes 4×4, 4×5, 5×4 and 5×5 into four different patterns. Based on variable $\bar{X}$, Control Chart equations on each window splitting; Upper Control Limit ($\text{UCL} = \bar{\bar{x}} + A_2\bar{R}$), Centre line ($\text{CL} = \bar{\bar{x}}$) and Lower Control Limit ($\text{LCL} = \bar{\bar{x}} - A_2\bar{R}$) where $\bar{R}$ is average range of samples; $\bar{\bar{x}}$ is population mean and A_2 [33] is constant, applied [28, 34] to derive threshold which is used on the image for segmentation.

4 Assessing Segmented Regions Through SSIM Index

Mean Structural SIMilarity (SSIM) measures the image quality by computing the similarity between images. SSIM [35] is designed to improve the traditional methods such as peak signal-to-noise ratio (PSNR) and mean squared error (MSE). The resultant SSIM index is a decimal value. SSIM is calculated between the two windows, F and G, of common window size $n \times n$. The common window size is set as 3×3 in this study. A higher value of SSIM indicates that the resultant image has more similarity to the original image. Also, the formulation in Eq. (14) given below satisfies a number of properties

(a) Guarantees ***symmetry***, meaning that $\text{SSIM}(F, G) = \text{SSIM}(G, F)$; thus also ensure post swapping on the original and coded images
(b) Ensures ***boundedness***, in the sense that $\text{SSIM}(F, G) < 1$.
(c) Unique ***maximum***, meaning that $\text{SSIM}(F, G) = 1$ if and only if $F = G$.

$$\text{SSIM}(F, G) = \frac{(2\mu_F\mu_G + C_1)(2\sigma_{FG} + C_2)}{\left(\mu_F^2 + \mu_G^2 + C_1\right)\left(\sigma_F^2 + \sigma_G^2 + C_2\right)} \quad (14)$$

where μ_F is the average of F; μ_G is the average of G; α_F^2 and α_G^2 are the variance of F and G respectively; σ_{FG} is the covariance; C_1 and C_2 are stabilizing parameters. Generally, $C_1 = 0.05$ and $C_2 = 0.05$. Recall that the covariance between F and G is defined as

$$\sigma_{FG} = \frac{1}{L-1}\sum_{i=0}^{L-1}(F_i - \mu_F)(G_i - \mu_G) \quad (15)$$

Their mean SSIM is defined as

$$\text{SSIM}(F, G) = \frac{1}{W}\sum_{j=1}^{W}\left(F_j G_j\right) \quad (16)$$

where W is the number of local windows of size 3×3 in the image.

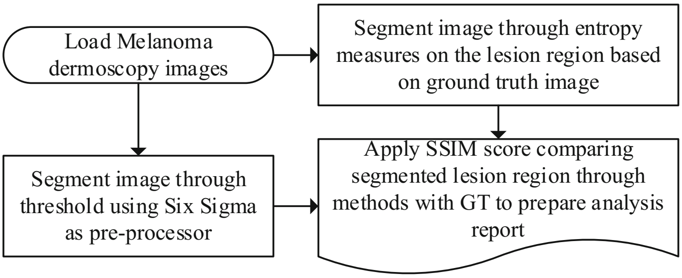


Fig. 1 Methodology for generating segmentation and assessment through SSIM

5 Tools and Methodology

For malignant melanoma, appropriate imaging of skin lesions is visioned as a priority role for diagnostic aid. In this proposed method uses digitized dermoscopy skin lesion images in RGB colour space with a similar size of 1024 × 768 pixels. The image set composed from 1636 digital dermoscopy images from by Dr. William V. Stoecker, Stoecker & Associates, USA. Images obtained using non-polarized contact dermoscopy techniques. These images may contain hair which was preprocessed using SharpRazor, a state-of-the-art tool. The methodology for generating results is explained in Fig. 1.

6 Results and Discussion

Segmented region through entropy and Six Sigma was assessed through SSIM, it is based on their score results with respect to image plane listed in Table 1. It was evidenced that for all the entropy measure their SSIM score falls for a maximum number

Table 1 Based on 1636 melanoma images samples and their related SSIM score on various segmented methods aligned on similar value ranges

SSIM score	Kapur	Harda and Charvat	Renyi	Tsallis	Six Sigma
Within >=0 and <=0.1	56	44	5	17	3
Within >0.1 and <=0.2	56	72	21	78	9
Within >0.2 and <=0.3	88	95	67	143	18
Within >0.3 and <=0.4	190	181	160	197	97
Within >0.4 and <=0.5	308	321	300	342	245
Within >0.5 and <=0.6	***371***	***359***	***384***	***354***	360
Within >0.6 and <=0.7	265	273	332	247	**396**
Within >0.7 and <=0.8	202	199	246	181	305
Within >0.8 and <=0.9	90	83	110	70	174
Within >0.9 and <=1	10	9	11	7	29

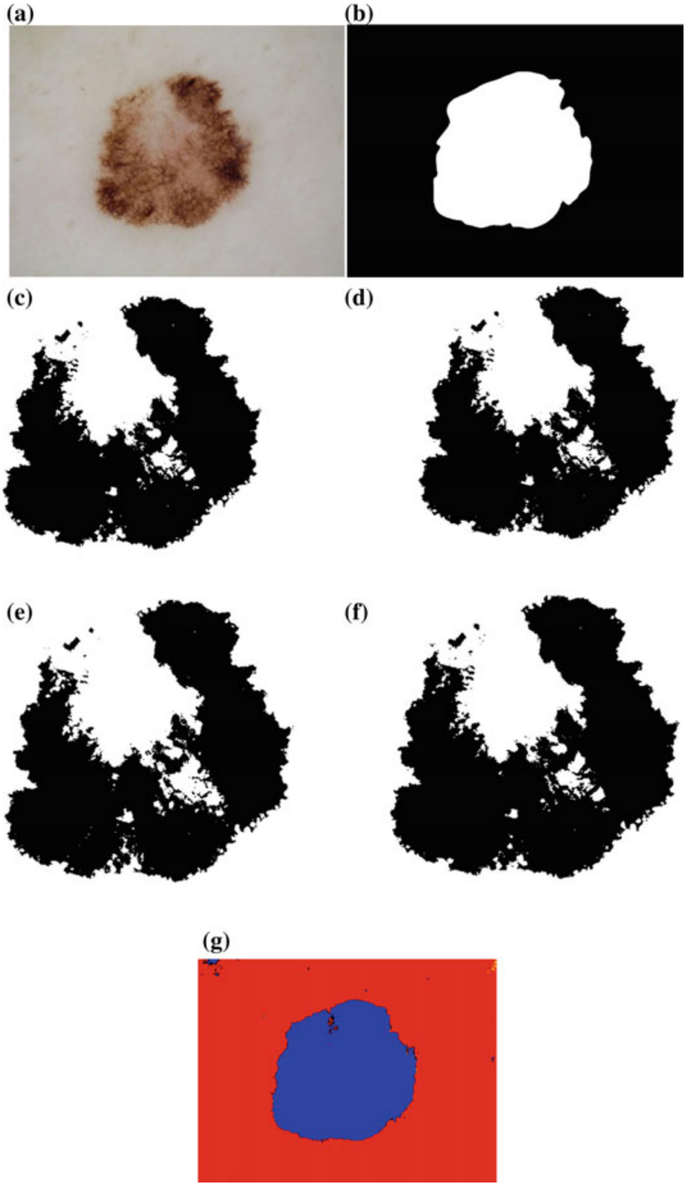


Fig. 2 Segmented image results on image post applying threshold through different methods. **a** Malignant melanoma dermoscopy image; **b** Manual marked lesion region (GT) image; segmented image through **c** Havrda and Charvat entropy; **d** Kapur entropy; **e** Renyi entropy **f** Tsallis entropy and **g** Segmented based on Six Sigma based threshold as preprocessor

of images within $\{> 0.5, \leq 0.6\}$. In case if Six Sigma score falls maximum within $\{> 0.6, \leq 0.7\}$ which shows that Six Sigma tries to give better optimized segmented region in comparison with Ground Truth (GT) image than the entropy measures. Figure 2 demonstrates sample malignant melanoma image, its expert identified GT and corresponding segmented results through proposed methods. Future work was planned to assess the applicability of various thresholds for detecting the internal borders with metrics for quantifying the spreading of disease. Results of various samples and the executable codes are made available at https://drive.google.com/open?id=1Z0a9frSZB2YImYHld3qvWi38QBHlNKNt.

Acknowledgements This publication was made possible by SBIR Grants R43 CA153927-01 and CA101639-02A2 of the National Institutes of Health (NIH). Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the NIH. Further, the authors would like to acknowledge the support rendered by the management of SASTRA Deemed to be University, Tirumalaisamudram, India.
This work is a by-product of generic image processing tool named "Bhadraloka" being developed (Dot Net platform) in SASTRA for the project titled "Development of techniques for processing radiographic images for automated detection of defects" with the funding assistance from Board of Research in Nuclear Science (BRNS), Department of Atomic Energy, Government of India (No. 2013/36/40-BRNS/2305).
Also, SS would like to thanks HCL Technologies Limited in supporting all through this research work.

References

1. Siegel RL, Miller KD, Jemal A (2017) Cancer statistics. CA Cancer J Clin 64(1):9–29
2. Weinstein DA, Konda S, Coldiron BM (2017) Use of skin biopsies among dermatologists. Dermatol Surg 43(11):1348–13657
3. Esteva A, Kuprel B, Novoa RA et al (2017) Dermatologist-level classification of skin cancer with deep neural networks. Nature 542(7639):115–118
4. Ferris LK, Harkes JA, Gilbert B et al (2015) Computer-aided classification of melanocytic lesions using dermoscopic images. J Am Acad Dermatol 73(5):769–76
5. Pehamberger H, Binder M, Steiner A et al (1993) In vivo epiluminescence microscopy: improvement of early diagnosis of melanoma. J Invest Dermatol 100:356S–362S
6. Soyer HP, Argenziano G, Chimenti S, Ruocco V (2001) Dermoscopy of pigmented skin lesions. Eur J Dermatol 11(3):270–277
7. Soyer HP, Argenziano G, Talamini R, Chimenti S (2001) Is dermoscopy useful for the diagnosis of melanoma? Arch Dermatol 137(10):1361–1363
8. Stolz W, Braun-Falco O, Bilek P, Landthaler M, Burgdorf WHC, Cognetta AB (eds) (2002) Color atlas of dermatoscopy. Wiley-Blackwell, Hoboken
9. Braun RP, Rabinovitz HS, Oliviero M, Kopf AW, Saurat JH (2002) Pattern analysis: a two-step procedure for the dermoscopic diagnosis of melanoma. Clin Dermatol 20(3):236–239
10. Boldrick JC, Layton CJ, Nguyen J, Swetter SM (2007) Evaluation of digital dermoscopy in a pigmented lesion clinic: clinician versus computer assessment of malignancy risk. J Am Acad Dermatol 56(3):417–421
11. Perrinaud A, Gaide O, French LE, Saurat JH, Marghoob AA, Braun RP (2007) Can automated dermoscopy image analysis instruments provide added benefit for the dermatologist? A study comparing the results of three systems. Br J Dermatol 157(5):926–933

12. Mishra NK, Celebi ME (2016) An overview of melanoma detection in dermoscopy images using image processing and machine learning. arXiv preprint arXiv:1601.07843
13. Rosendahl et al (2012) Dermatoscopy in routine practice: 'Chaos and clues'. Aust Fam Physician 41(7):482
14. Friedman RJ et al (1985) Early detection of malignant melanoma: the role of physician examination and self-examination of the skin. CA Cancer J Clin 35(3):130–151
15. Rubegni P et al (2015) Computer-assisted melanoma diagnosis: a new integrated system. Melanoma Res 25(6):537–542
16. Andreassi L et al (1999) Digital dermoscopy analysis for the differentiation of atypical nevi and early melanoma: a new quantitative semiology. Arch Dermatol 135(12):1459–1465
17. Landau M et al (1999) Computerized system to enhance the clinical diagnosis of pigmented cutaneous malignancies. Int J Dermatol 38(6):443–446
18. Umbaugh SE et al (1989) Automatic color segmentation of images with application to detection of variegated coloring in skin tumors. Eng Med Biol Mag IEEE 8(4):43–50
19. Aitken JF et al (1996) Reliability of computer image analysis of pigmented skin lesions of Australian adolescents. Cancer 78(2):252–257
20. Ercal F et al (1994) Neural network diagnosis of malignant melanoma from color images. IEEE Trans Biomed Eng 41(9):837–845
21. Ganster H et al (2001) Automated melanoma recognition. IEEE Trans Med Imaging 20(3):233–239
22. Kaushik RHC et al. (2013) The median split algorithm for detection of critical melanoma color features. In: International conference on computer vision theory and applications (VISAPP), pp 492–495
23. Stanley RJ et al (2007) A relative color approach to color discrimination for malignant melanoma detection in dermoscopy images. Skin Res Technol 13(1):62–72
24. Almubarak HA et al (2017) Fuzzy color clustering for melanoma diagnosis in dermoscopy images. Information 8(3):89
25. Sabbaghi Mahmouei SA et al. (2015) An improved colour detection method in skin lesions using colour enhancement. In: Australian biomedical engineering conference (ABEC 2015)
26. Madooei A et al (2013) A colour palette for automatic detection of blue-white veil. In: Color and imaging conference, vol 2013, no 1, pp 200–205
27. Tiwari R, Sharma B (2016) A comparative study of Otsu and entropy based segmentation approaches for lesion extraction. In: Conference: 2016 international conference on inventive computation technologies (ICICT)
28. Sankaran S, Malarvel M, Sethumadhavan G, Sahal D (2017) Quantitation of malarial parasitemia in giemsa stained thin blood smears using six sigma threshold as preprocessor. Optik Int J Light Electr Opt 145:225–239. ISSN 0030-4026, http://dx.doi.org/10.1016/j.ijleo.2017.07.047
29. Comparison of Shannon, Renyi and Tsallis Entropy used in Decision Trees, Tomasz Maszczyk and Wlodzislaw Duch
30. Ja Havrda, František Charvát (1967) Quantification method of classification processes: concept of structural a-entropy. Kybernetika 03(1):30–35
31. Kapur JN, Sahoo PK, Wong AKC (1985) A new method for Gray-level picture thresholding using the entropy of the histogram. Comp Vis Graphics Image Process 29:273–285. https://doi.org/10.1016/0734-189X(85)90125-2
32. Rényi A (1961) On measures of entropy and information. In: Proceedings of the fourth Berkeley symposium on mathematical statistics and probability, Volume 1: contributions to the theory of statistics, pp 547–561. University of California Press, Berkeley, California. http://projecteuclid.org/euclid.bsmsp/1200512181
33. http://web.mit.edu/2.810/www/files/readings/ControlChartConstantsAndFormulae.pdf
34. Sankaran S, Sethumadhavan G (2013) Quantifications of asymmetries on the spectral bands of MALIGNANT melanoma using six sigma threshold as preprocessor. In: Third international conference on computational intelligence and information technology (CIIT 2013), Mumbai, pp 80–86. https://doi.org/10.1049/cp.2013.2575

35. Wang Z, Bovik AC, Sheikh HR, Simoncelli EP (2004) Image quality assessment: from error visibility to structural similarity. IEEE Trans Image Process 13(4):600–612. https://doi.org/10.1109/tip.2003.819861

Impact of Speckle Filtering on the Decomposition and Classification of Fully Polarimetric RADARSAT-2 Data

Sivasubramanyam Medasani and G. Umamaheswara Reddy

Abstract Decomposition and classification are vital processing stages in polarimetric synthetic aperture radar (PolSAR) information processing. Speckle noise affects SAR data since backscattered signals from various targets are coherently integrated. Current study investigated the impact of speckle suppression on the target decomposition and classification of RADARSAT-2 fully polarimetric data. Speckle filters should suppress the speckle noise along with the retention of spatial and polarimetric information. The performance of improved Lee–Sigma, intensity-driven adaptive neighborhood (IDAN), refined Lee, and boxcar filters were assessed utilizing the spaceborne dataset, that is, fully polarimetric RADARSAT-2 C-band SAR data for the Mumbai region, India. The effect of speckle suppression on target decomposition was analyzed in this study. Different speckle noise suppression techniques were applied to RADARSAT-2 dataset, followed by Yamaguchi three-component and VanZyl decompositions. The obtained findings revealed that the improved Lee–Sigma filter demonstrated better volume scatterings in forest areas and double bounce in urban areas than the other techniques considered in the analysis. Additionally, the efficacy of the different speckle suppression techniques listed above was assessed. The effectiveness of the speckle filtering algorithm was evaluated by applying the Wishart supervised classification to the filtered and unfiltered data. IDAN, boxcar, refined Lee, and improved Lee–Sigma filters were assessed to find the classification accuracy improvement. A considerable amount of improvement was observed in the classification accuracy for mangrove and forest classes. Minimal enhancement was detected for settlement, bare soil, and water classes.

S. Medasani (✉) · G. Umamaheswara Reddy
Department of Electronics and Communication Engineering, Sri Venkateswara University College of Engineering, Sri Venkateswara University, Tirupati, Andhra Pradesh, India
e-mail: medasani7@gmail.com

G. Umamaheswara Reddy
e-mail: umaskit@gmail.com

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_20

1 Introduction

Nowadays SAR imaging, a coherent technique for microwave remote sensing, is used for producing large-scale two-dimensional surface reflectivity images of the Earth; these images are of high resolution. Solar illumination does not influence the performance of microwave remote sensing sensors because of the active operation mode; thus, images can be taken during the day or night. Furthermore, the effects of smoke, fog, clouds, and rain can be avoided by working in the microwave frequencies. Emerging polarimetric SAR (PolSAR) [1, 2] images increase the applicability of SAR remote sensing and provide an additional dimension to the SAR information. PolSAR images have various applications including obtaining information of soil moisture; determining surface roughness, mapping snow and ice; and classifying land cover. These obtained factors are used in various areas including forestry, urbanism, agriculture, industry, and geology.

Currently, using polarimetry for radar remote sensing has achieved substantial attention. The use of detected microwave scattering by volume and surface structures to extract the physical information is essential for radar remote sensing applications. The coherency matrix $[T]$ of dimension 3×3 is the most crucial matrix measured by polarimetric SAR systems. This matrix illustrates the local differences observed in the scattering matrix. Moreover, an operator of the lowest order is appropriate for extracting polarimetric parameters of distributed scatterers when either additive or multiplicative noise or both are present. In radar remote sensing, a multiplicative statistical description is needed by several targets of concern because of the combined random vector scattering effects and coherent speckle noise from volume and surface. To achieve the classification or inversion of scatter data of such targets, a scattering mechanism that is dominant or average should be generated. The aforementioned averaging procedure engenders the notion of distributed targets, which exhibit a unique structure, unlike pure single or stationary targets.

The most crucial application of PolSAR is the terrain and land-use classification. For the supervised and unsupervised classification of terrains, several algorithms are developed. Moreover, for the supervised classification of terrains, training sets were selected for each class on the basis of scattering contrast differences or ground truth maps observed in PolSAR images. The PolSAR response for all images was ingrained in three complex and three real parameters—total nine parameters. Because of the high dimensions of PolSAR data, the selection of training sets may become difficult in case the ground truth maps are unavailable. In contrast, the classification of the unsupervised terrain involves the automatic classification of images by finding clusters on the basis of a specific criterion. Classification methods based on information obtained from the scattering mechanisms include the support vector machine, eigenvalue-based or model-based decompositions, neural networks, and maximum likelihood algorithm with the complex Wishart distribution of the coherency or covariance matrix. This study examined the effect that speckle suppression has on the target decomposition and classification of fully polarimetric RADARSAT-2 data.

Table 1 Specifications of RADARSAT-2

Frequency band	Mode	Incidence angle (°)	Date of acquisition	Polarization	Slant range resolution (m)
C	FQ22 (Ascending right)	41.7	February, 16 2011	Quad Pole	8.0

FQ fine quad

Section 2 provides RADARSAT-2 specifications. Section 3 discusses the procedure of the proposed work; Sect. 4 explains the results. At the end of this paper, Sect. 5 provides the overall conclusions.

2 Dataset and Study Area

The spaceborne dataset considered in the current study constitutes fully polarimetric RADARSAT-2 C-band SAR data obtained for Mumbai. Specifications of the RADARSAT-2 sensor are given in Table 1.

3 Methodology

3.1 Assessment of Speckle Suppression Techniques by Using Fully Polarimetric RADARSAT- 2 Data

Speckle noise affects SAR information since backscattered signals from various targets are coherently integrated. For visual and quantitative evaluation of speckle filters, refined Lee, boxcar, intensity-driven adaptive neighborhood (IDAN), and improved Lee–Sigma filters were applied to the fully polarimetric RADARSAT-2 C-band SAR data (Fig. 1). The aforementioned filters were quantitatively compared using various parameters [3–5] such as the edge preservation index (EPI), mean square error (MSE), speckle suppression index (SSI), peak signal-to-noise ratio (PSNR), bias, standard deviation to mean ratio (SD/M), and mean and standard deviations of the ratio image.

3.2 Impact of Speckle Suppression on the Target Decomposition

Target decomposition techniques [1, 6] are classified on the basis of the following factors: (i) the coherent decomposition of the scattering matrix (Krogager and

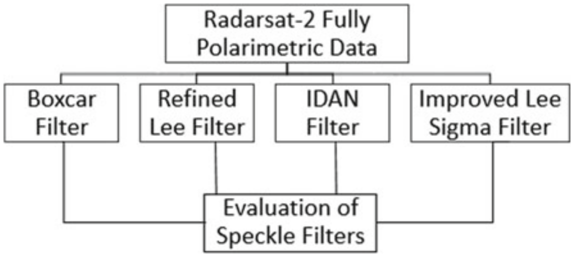


Fig. 1 Assessement of various speckle suppression techniques

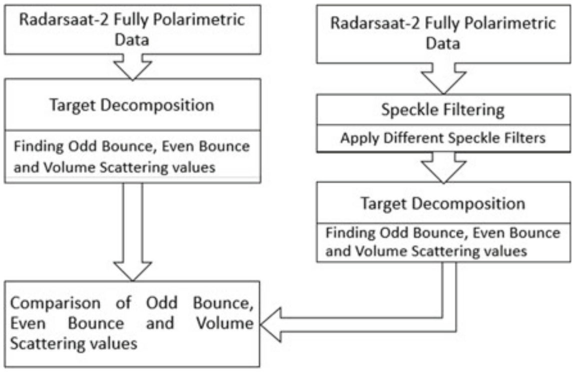


Fig. 2 Impact of speckle suppression on the target decomposition

Cameron), (ii) the dichotomy of the Kennaugh matrix (Huynen and Barens), (iii) the model-based decomposition of the coherency or covariance matrix (Freeman and Durden, and Dong), and (iv) the analysis of eigenvalue or eigenvector of the coherency or covariance matrix (Cloude, VanZyl, Cloude, and Pottier). In this study, the Yamaguchi three-component [7, 8] and VanZyl decompositions [9] were applied for analyzing the speckle filtering impact on decomposition. In [10] various speckle suppression techniques are evaluated using hybrid polarimetric data (RISAT-1) and also the impact of speckle suppression on the target decomposition of hybrid polarimetric data is presented. Figure 2 presents the block diagram of the method employed to determine the influence of speckle suppression on the target decomposition of fully polarimetric RADARSAT-2 data.

3.3 Impact of Speckle Suppression on Classification

The most crucial application of PolSAR is the land-use and terrain classification. Several algorithms have been generated for the unsupervised and supervised terrain classification. The PolSAR classifications, both unsupervised and supervised, are generally influenced by averaging or filtering processes such as the multi-look technique that is employed to reduce speckle noise. The Wishart classification [11–13] is a renowned technique utilized to obtain the supervised classification of PolSAR

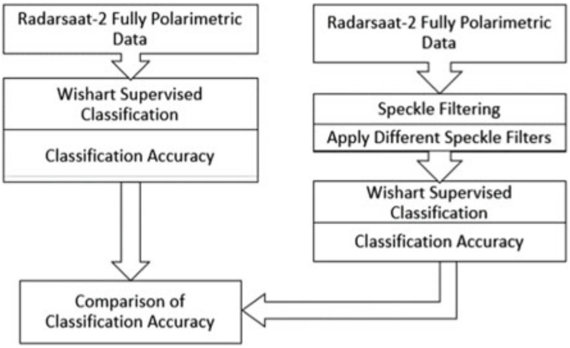


Fig. 3 Impact of speckle suppression on classification

image data. The size of the sliding window size [14] for boxcar and refined Lee filters is a crucial parameter for speckle suppression. Speckle filters are evaluated using hybrid polarimetric data (RISAT-1) and also the impact of speckle filtering on the classification of hybrid polarimetric data is presented in [10]. The proposed flowchart for the comparison of the Wishart supervised classification is displayed in Fig. 3.

4 Results and Discussion

4.1 Assessment of Speckle Suppression Techniques by Using Fully Polarimetric RADARSAT- 2 Data

For the visual and quantitative evaluation of the speckle filters, various aforementioned filters were applied to fully polarimetric data for Mumbai. Figure 4 displays the results of RADARSAT-2 C-band data.

After the visual examination, the filtered images were quantitatively compared. Various parameters were used for the quantitative comparison including MSE, PSNR, EPI, SD/M, bias, SSI, and the standard deviation and mean of the ratio image. The highest values of PSNR and EPI represent optimal performance of the filters. Moreover, the lowest values of MSE, SD/M, bias, and SSI indicate optimal performance of the filters. The standard deviation and mean of the ratio image should be as small as possible and approximately equal to one, respectively.

The visual interpretation indicates that the boxcar filter degrades the spatial resolution because it does not include the internal mechanism for preserving the spatial resolution or details. The spatial resolution preservation for the refined Lee filter exhibited improvement; however, a considerable smoothing effect was still remarked in the filtered data. IDAN filter uses adaptive neighborhoods as a spatial support, and

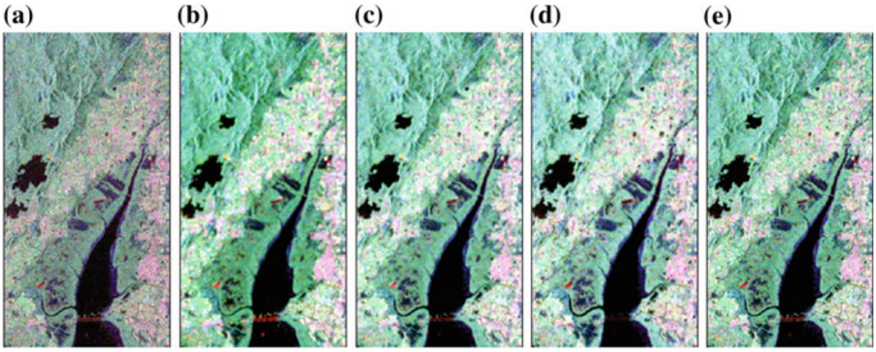


Fig. 4 Assessment of various speckle suppression techniques by using RADARSAT-2 data for the Mumbai region. **a** Original Pauli RGB and **b** boxcar, **c** refined Lee, **d** IDAN, and **e** improved Lee–Sigma filters

Table 2 Assessment of speckle suppression techniques (7 × 7 window) by using RADARSAT-2 data for the Mumbai region

Filter/Parameter	MSE	PSNR	EPI	SD/M	Bias	SSI
Boxcar	208.6176	24.9373	0.1078	0.4149	0.0570	0.8115
Refined Lee	203.6277	25.0424	0.1989	**0.4139**	0.0575	**0.8095**
IDAN	202.4071	25.0685	**0.2190**	0.4154	0.0592	0.8124
Improved Lee–Sigma	**198.8244**	**25.1461**	0.2072	0.4192	**0.0544**	0.8199

Bold text indicates better value or performance

these neighborhoods were derived from the intensity information. Due to the use of adaptive neighborhoods, the speckle noise was substantially reduced. However, the contour and fine details were preserved, thus avoiding the blurring effect. The improved Lee–Sigma filter, where the estimate of the minimum mean square error (MMSE) was used as a priori mean, exhibited an excellent preservation of fine details and strong targets.

The quantitative comparison of the various aforementioned filters employed in this study by using RADARSAT-2 data for the Mumbai region is given in Tables 2 and 3. Table 2 illustrates that the improved Lee–Sigma filter exhibited an improved performance in terms of MSE with the lowest value of 198.8244, PSNR with the highest value of 25.1461, and bias with the lowest value of 0.0544. The IDAN filter had superior performance in terms of EPI and had the highest value of 0.2190 than the other filters considered in this study. SD/M (0.4139) and SSI (0.8095) for the refined Lee filter were the lowest, respectively, compared with the other filters employed. For the ratio image, Table 3 displays that the boxcar filter exhibited the mean value of 0.9546 (~1), and the standard deviation (0.5963) for the improved Lee–Sigma filter was the lowest.

Table 3 Assessment of speckle suppression techniques (7 × 7 window) by using RADARSAT-2 data for the Mumbai region and the ratio image

Filter/Parameter	Mean	Standard Deviation
Boxcar	**0.9546**	0.9133
Refined Lee	0.7994	0.6813
IDAN	0.6891	0.7966
Improved Lee–Sigma	0.8031	**0.5963**

Bold text indicates superior value or performance

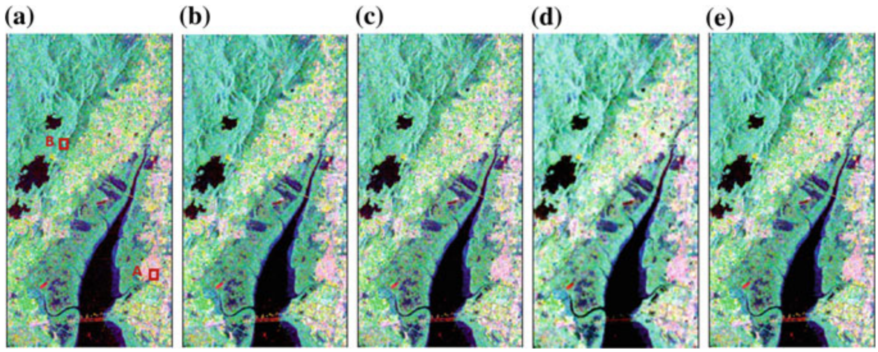


Fig. 5 Yamaguchi three-component decomposition images before and after employing the various speckle suppression techniques on RADARSAT-2 data for the Mumbai region. **a** Original data and **b** boxcar, **c** refined Lee, **d** IDAN, and **e** improved Lee–Sigma filters

4.2 Impact of Speckle Suppression on the Target Decomposition

Current study examined the influence of despeckling on the target decomposition. The speckle filters, including refined Lee, boxcar, improved Lee–Sigma, and IDAN filters, were applied to the RADARSAT-2 dataset. The Yamaguchi three-component and VanZyl decompositions were then conducted on the dataset. Figures 5 and 6, respectively, display the Yamaguchi three-component and VanZyl decomposition images before and after applying the various speckle suppression techniques on RADARSAT-2 data for the Mumbai region. In Figs. 7 and 8, the percentage of volume, even bounce, and odd bounce scatterings are displayed, respectively, by conducting the Yamaguchi three-component and VanZyl decompositions for RADARSAT-2 C-band data in the (A) urban and (B) forest areas, which are shown in Figs. 5a and 6a before and after employing the various speckle suppression techniques.

The variation in volume, single bounce, double-bounce scatterings of data for the urban area is displayed in Figs. 7a and 8a before and after employing the various speckle suppression techniques to conduct the Yamaguchi three-component and VanZyl decompositions, respectively, for RADARSAT-2 data. Double-bounce

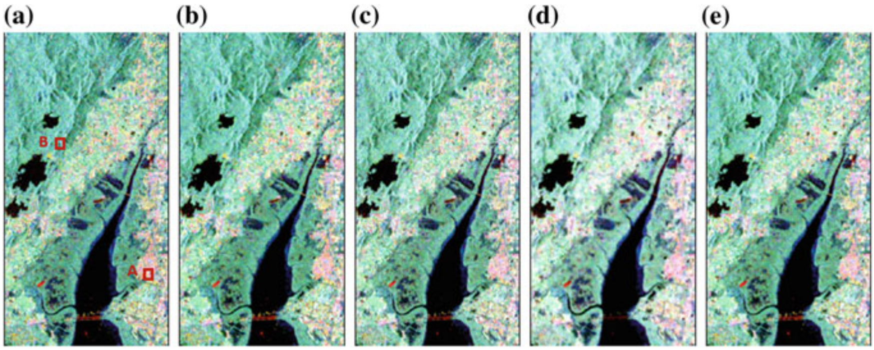
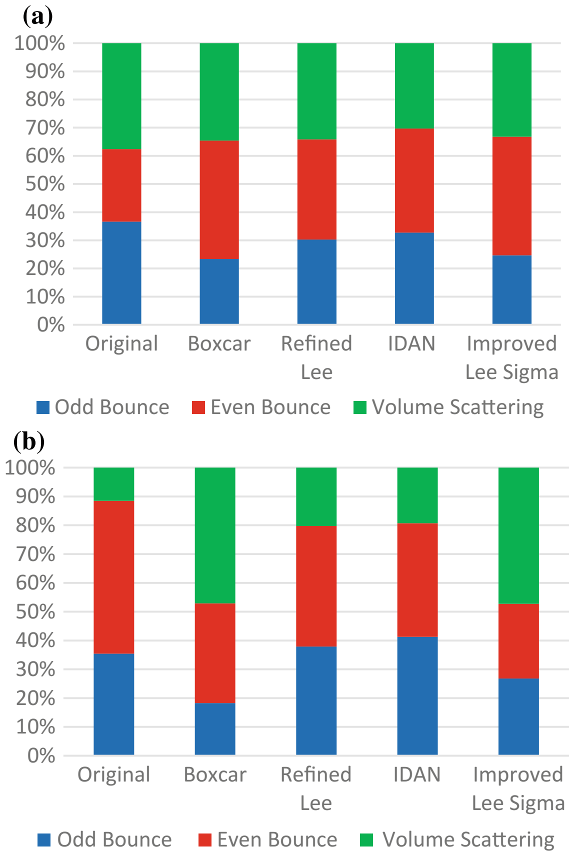


Fig. 6 VanZyl decomposition images before and after employing the various speckle suppression techniques on RADARSAT-2 data for Mumbai region. **a** Original data and **b** boxcar, **c** refined Lee, **d** IDAN, and **e** improved Lee–Sigma filters

Fig. 7 Variation in the volume, double bounce, and single bounce scatterings by conducting the Yamaguchi three-component decomposition before and after applying the various speckle filters on RADARSAT-2 data for the Mumbai region in the **a** urban and **b** forest areas



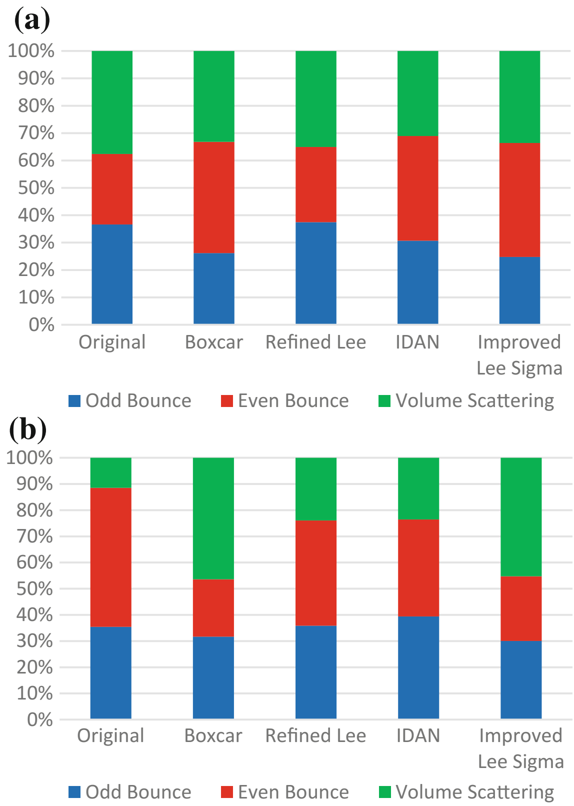


Fig. 8 Variation in volume, single bounce, and double-bounce scatterings by conducting VanZyl decomposition before and after employing the various speckle suppression techniques on RADARSAT-2 data for the Mumbai region in the **a** urban and **b** forest areas

scattering for all filters considered in this study for RADARSAT-2 dataset was considerably developed. However, for the boxcar filter, the edges exhibited more blurring than the other filters considered in this study because of indiscriminate averaging. The edge preservation was better in IDAN and improved Lee–Sigma filters than that in the refined Lee filter. The variation in double-bounce scattering in the refined Lee filter was less than that in the IDAN filter. Double-bounce scattering for the improved Lee–Sigma filter improved marginally compared with the IDAN filter. Figures 7b and 8b show the variation in volume, double bounce, and single bounce scatterings in the forest area before and after the application of different speckle suppression techniques for RADARSAT-2 data. All filters considered in the study exhibited improved volume scattering. However, refined Lee, IDAN, and improved Lee–Sigma filters exhibited better edge preservation than that of the boxcar filter.

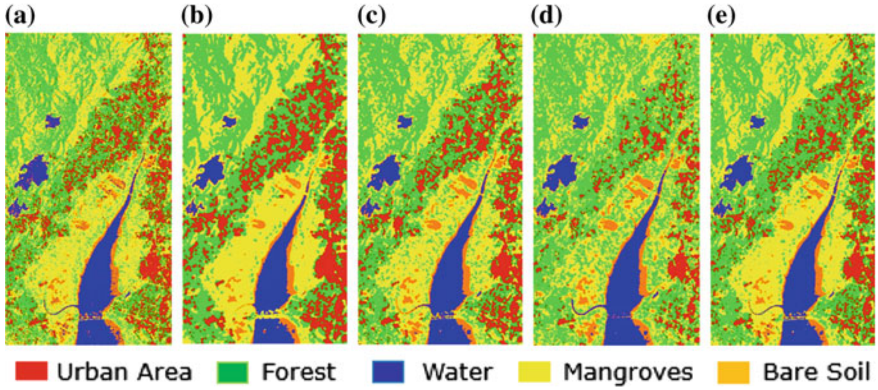


Fig. 9 Classification images of RADARSAT-2 data for the Mumbai region. **a** Original data and **b** boxcar, **c** refined Lee, **d** IDAN, and **e** improved Lee–Sigma filters

4.3 Impact of Speckle Suppression on Classification

An improvement in the accuracy of the classification due to the speckle suppression of fully polarimetric data is the essential target of current study. In this study, the impact of speckle suppression on classification accuracy was examined utilizing the fully polarimetric data. Various speckle suppression techniques—IDAN, boxcar, refined Lee, and improved Lee–Sigma were employed on fully polarimetric RADARSAT-2 data, followed by a Wishart supervised classification.

Figure 9 displays the Wishart supervised classification images of RADARSAT-2 C-band data for the Mumbai region before and after employing different speckle suppression techniques. For refined Lee, boxcar, and improved Lee–Sigma filters, the window size was 7 × 7. Table 4 indicates the variation in the classification accuracy before and after employing the various speckle suppression techniques for the quantitative comparison of different classes and the overall accuracy. Speckle filters considered in this study exhibited improved classification accuracy of the classes of the unfiltered data. The extent of improvement was different for all classes. The classification accuracy of several classes depends on the speckle suppression technique employed. RADARSAT-2 C-band data for the Mumbai region exhibited higher improvement in the classification accuracy for the forest and mangrove classes than the water, bare soil, and settlement classes.

For refined Lee and boxcar filters, the sliding window size is a essential parameter for speckle suppression. The result obtained by changing the window size was assessed for the refined Lee and boxcar filters by utilizing RADARSAT-2 C-band data. The classification accuracy of the speckle suppressed data was evaluated by changing the size of the window from 3 × 3 to 11 × 11. The effects of boxcar and refined Lee filters for the moving window sizes of several classes considered in the study are given in Tables 5 and 6, respectively, for RADARSAT-2 data.

Table 4 Assessment of the classification accuracies of various speckle suppression techniques (7 × 7 window) by using RADARSAT-2 data for the Mumbai region

Class	No filter	Boxcar	Refined Lee	IDAN	Improved Lee–Sigma
Settlement	98.21	100	100	100	100
Forest	63.33	69.41	69.41	67.92	70.45
Water	96.44	100	100	100	100
Mangrove	88.83	99.19	89.82	93.56	94.83
Bare soil	97.55	98.57	100	99.80	99.18
Overall accuracy	88.43	93.79	92.45	92.37	93.03

Table 5 Assessment of classification accuracies of the boxcar filter with various window sizes by using RADARSAT-2 data for the Mumbai region

Class	No Filter	3 × 3	5 × 5	7 × 7	9 × 9	11 × 11
Settlement	98.21	100	100	100	100	100
Forest	63.33	62.87	64.98	69.41	74.68	79.54
Water	96.44	99.42	99.81	100	100	100
Mangrove	88.83	91.76	95.15	99.19	99.03	99.35
Bare soil	97.55	97.34	97.75	98.57	98.57	97.95
Overall accuracy	88.43	90.11	91.73	93.79	94.74	95.61

Table 6 Assessment of classification accuracies of the refined Lee filter with various window sizes by using RADARSAT-2 data for the Mumbai region

Class	No filter	3 × 3	5 × 5	7 × 7	9 × 9	11 × 11
Settlement	98.21	98.74	99.08	100	100	100
Forest	63.33	61.81	64.14	69.41	73.93	76.35
Water	96.44	98.45	99.25	100	100	100
Mangrove	88.83	90.11	91.18	92.82	95.64	98.71
Bare soil	97.55	96.23	98.77	100	100	100
Overall accuracy	88.43	89.15	90.30	92.45	93.93	95.12

Tables 5 and 6 indicate that the refined Lee filter requires a window of a larger size for obtaining an accuracy level identical to that of the boxcar filter. However, the edge preservation of the refined Lee filter was better than the boxcar filter. The percentage improvement of the relative classification accuracies for RADARSAT-2 C-band data by using the refined Lee and boxcar filters on the unfiltered data are given in Table 7 for the window size of 7 × 7. The assessment indicates a considerable amount of enhancement in the accuracy of the classification for the mangrove and forest classes. A minimal enhancement was seen for the water, bare soil, and settlement classes.

Table 7 Percentage of relative improvement in the classification accuracy for the original data by using the boxcar and refined Lee speckle filters

Filter/Class	Settlement	Forest	Water	Mangrove	Bare Soil
Boxcar	2	10	4	12	1
Refined Lee	2	10	4	5	3

5 Conclusions

Fully polarimetric RADARSAT-2 data was used to evaluate IDAN, boxcar, improved Lee–Sigma, and refined Lee filters. The boxcar filter degrades the spatial resolution because it does not comprise the internal mechanism for preserving spatial resolution or details. The refined Lee filter improved the spatial resolution preservation; however, the filtered data was considerably smooth. The spatial strategy was improved using the IDAN filter; this filter satisfactorily preserved the spatial resolution and details, although some data were evidently smoothed. The computational efficiency of improved Lee–Sigma and refined Lee filters should be similar because both algorithms are based on local statistics. The improved Lee–Sigma filter, where the estimate of MMSE is used as the priori mean, exhibited an excellent preservation of fine details and strong targets.

In the current study, the impact of speckle suppression on the target decomposition was assessed using fully polarimetric RADARSAT-2 data. A significant amount of enhancement was observed in volume and double-bounce scatterings in forest and urban areas for all the speckle suppression techniques deliberated in this study. The speckle reduction impact on the classification accuracy was assessed using fully polarimetric data. The speckle suppression techniques considered in this study exhibited enhancement in the classification accuracy of classes of the unfiltered data. The extent of improvement was different for all classes. The classification accuracy of various classes depends on the filtering technique employed. RADARSAT-2 C-band data for the Mumbai region exhibited higher enhancement in the classification accuracy for the mangrove and forest classes than that for the water, bare soil, and settlement classes. For refined Lee and boxcar filters, the moving window size is an essential parameter for speckle suppression. RADARSAT-2 C-band data was used to evaluate the result of the changing sliding window size for the refined Lee and boxcar filters. The classification accuracy of data with suppressed speckle was evaluated by varying the size of the window from 3×3 to 11×11. The refined Lee filter requires a window with a larger size for achieving classification accuracy identical to that of the boxcar filter. However, the edge preservation performance of the refined Lee filter is better than that of the boxcar filter.

Acknowledgements The authors are grateful to Space Application Centre, ISRO, India for giving the opportunity to carry out research work and providing the data under TREES. The authors are thankful to Dr. Anup Kumar Das, SAC, ISRO for providing the guidance to conduct the research. The authors are thankful to Dr. C. V. Rao, NRSC, ISRO for his constant support and encouragement. The authors are grateful to the Centre of Excellence and Department of Electronics and Communication

Engineering at Sri Venkateswara University College of Engineering for providing the resources. Furthermore, the authors are thankful to Mr. P. Anil Kumar, Mr. C. Raju, Mr. N. Chintaiah, and research scholars for the valued discussions and encouragement. The authors would like to thank the European Space Agency for providing the open-source software and the experimental data of the PolSARpro project.

References

1. Lee JS, Pottier E (2009) Polarimetric radar imaging: from basics to applications. CRC Press, Cleveland
2. Cloude SR (2009) Polarisation applications in remote sensing. Oxford University Press, Oxford
3. Foucher S, López-Martínez C (2014) Analysis, evaluation, and comparison of polarimetric SAR speckle filtering techniques. IEEE Trans Image Process 23(4):1751–1764
4. Argenti F, Lapini A, Alparone L, Bianchi T (2013) A tutorial on speckle reduction in synthetic aperture radar images. IEEE Geosci Remote Sens Mag 1:6–35
5. Di Martino G, Poderico M, Poggi G, Riccio D, Verdoliva L (2014) Benchmarking framework for SAR despeckling. IEEE Trans Geosci Remote Sens 52(3):1596
6. Cloude SR, Pottier E (1996) A review of target decomposition theorems in radar polarimetry. IEEE Trans Geosci Remote Sens 34(2):498–518
7. Freeman A, Durden S (1998) A three-component scattering model for polarimetric SAR data. IEEE Trans Geosci Remote Sens 36(3):963–973
8. Yamaguchi Y, Moriyama T, Ishido M, Yamada H (2005) Fourcomponent scattering model for polarimetric SAR image decomposition. IEEE Trans Geosci Remote Sens 43(8):1699–1706
9. Van Zyl JJ (1992) Application of Cloude's target decomposition theorem to polarimetric imaging radar data. In: Proceedings SPIE conference on radar polarimetry, San Diego, CA, vol 1748, pp 184–212
10. Medasani S, Umamaheswara Reddy G (2018) Speckle filtering and its influence on the decomposition and classification of hybrid polarimetric data of RISAT-1. Remote Sens Appl: Environ Soc 10:1–6
11. Lee JS, Grunes MR, Kwok R (1994) Classification of multi-look polarimetric SAR imagery based on complex Wishart distribution. Int J Remote Sens 15(11):2299–2311
12. Lee JS, Grunes MR, Ainsworth TL, Li-Jen D, Schuler DL, Cloude SR (1999) Unsupervised classification using polarimetric decomposition and the complex Wishart classifier. IEEE Trans Geosci Remote Sens 37(5):2249–2258
13. Ferro-Famil L, Pottier E, Lee JS (2001) Unsupervised classification of multifrequency and fully polarimetric SAR images based on the H/A/Alpha-Wishart classifier. IEEE Trans Geosci Remote Sens 39(11):2332–2342
14. Shitole S, De S, Rao YS, Mohan BK, Das A (2015) Selection of suitable window size for speckle reduction and deblurring using SOFM in polarimetric SAR images. J Indian Soc Remote Sens 43(4):739–750

Estimation of Precipitation from the Doppler Weather Radar Images

P. Anil Kumar and B. Anuradha

Abstract Estimating the rainfall from Radar observation plays an important role in the hydrological research. The Radar Rainfall plays a fundamental role in weather modeling and forecasting applications. Doppler Weather Radar (DWR) is used estimating the rainfall within 120 km from the Radar station. Rainfall intensity data obtained from the Surface Rainfall Intensity (SRI) product of DWR has been validated with the rain gauges located at Automatic Weather Station (AWS) data. Image processing methods such as edge detection and color identification are used to extract the rainfall from the SRI product. Time series rainfall over a particular location is compared with the AWS data using statistical parameters like correlation coefficient and Squared Pearson coefficient. The experimental results convey that the proposed method yields the high amount of accuracy. Graphical User Interface is developed to extract the point rainfall and time series rainfall over different locations within the range of Radar.

1 Introduction

Rain Gauges are used to measure the rainfall at different points in the field, the spatial variation of rainfall cannot be measured due to lack of sufficient rain gauges. In view of the practical problem of installing rain gauges at different stations, Radar has found to be an alternative. Radar has been used in predicting the weather over the last four to five decades. Radar can provide information such as rain, reflectivity, wind speed, and direction in both time and space varying domains. The underlying principle of radar is the Rayleigh Scattering in which scattering of electromagnetic radiation occurs by particles smaller than the wavelength of the radiation [1]. The meteorological radars are characterized by their operating electromagnetic frequencies such as X, C, or S bands. The weather Radars look deeper into the system to provide information such as Intensity, Rain Rate, Vertical Extent, and Drop Size Distribution. The Primary

P. Anil Kumar (✉) · B. Anuradha
Department of ECE, SVU College of Engineering, SV University, Tirupati, India
e-mail: anilkumar417@gmail.com

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_21

products of the DWR are Reflectivity (Z), Radial Velocity (V), and Spectrum Width (W). The Secondary products derived from Primary products such as Surface Rainfall Intensity (SRI) and Precipitation Accumulation (PAC) products are among them.

Radar Rainfall data derived from S-band DWR was used as input to hydrological modeling. Hydrologic Engineering Center-Hydrological modeling System (HEC-HMS) simulation results are calibrated and validated with the observed data [2]. Raindrop scatters pulses back to the receiver by the process of reflection, refraction, and scattering. The reflected signal strength and the time delay between the pulses define the distance of the target. Rainfall intensity is estimated from the reflectivity with the help of Marshall–Palmer Equation as shown in Eq. 1. The value of the constants decides the type of geographic area and precipitation such as convective rain, thunderstorm, orographic rain, stratiform and snow. The parameters for the Marshall–Palmer are fine-tuned for different geographical conditions and different climatic conditions [3, 4]. The values of a and b are 267 and 1.345 for the secondary product SRI of S-band DWR Chennai. The Reflectivity information was extracted from the DWR MAX (Z) product using image processing techniques. The time series Reflectivity extracted over a particular location was plotted and cloud vertical profile was obtained [5].

$$Z = N(D)D^6 = a\,R^b \tag{1}$$

where

Z	is the measured radar reflectivity in mm^6/m^3
D	is the drop diameter in mm
$N(D)$	is the number of drops of given diameter per cubic meter
R	is the rainfall rate at the ground level in mm/h
a, b	are constants which depend on latitude, longitude, season, and types of rainfall.

2 Data

The DWR SRI images from the Indian Meteorological Department (IMD) located at Chennai were downloaded continuously for every 10 min since November 2015 and a database is maintained (http://www.imd.gov.in). The AWS Data for the Chennai and surrounding locations were collected.

3 Methodology

The DWR SRI image which contains the convective activity within the range of DWR is shown in Fig. 1. The resolution of DWR SRI is 0.4 km/pixel. It covers a range

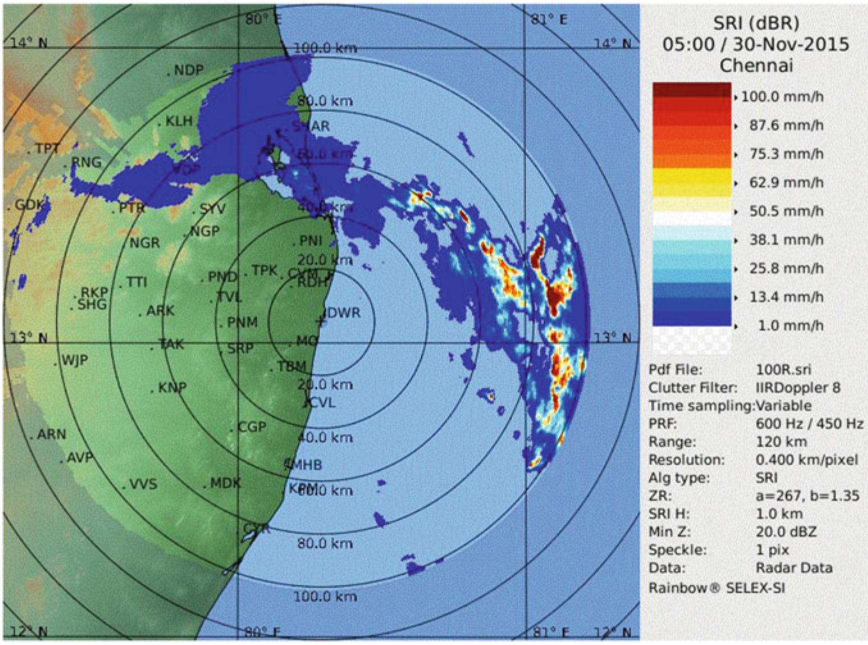


Fig. 1 Convective SRI image

of 120 km from the radar center. The minimum detectable reflectivity for rainfall calculation is 20 dBZ for the SRI product. The SRI product displays the value of rainfall intensity over each pixel of the cylindrical volume in its plane view. The color bar provided on the right side of the image gives information about the intensity of precipitation over a certain area of the image. The blue color corresponds to 1 mm/h and red color indicates the rainfall intensity of 100 mm/h. Similarly, each color in the color bar signifies the different intensity level of precipitation. The flowchart for rainfall extraction from SRI image is shown in Fig. 2.

An SRI image with zero convective activity at all locations is considered as a reference image (I_{ref}). An image which contains convective activity or the image which is of interest is considered as an image of interest (I_{int}). The images which are downloaded from the IMD website contains salt and pepper noise. Median filter which has the characteristics of scale-invariant nonlinear smoother is used to eliminate the noise by preserving the edges. Image preprocessing is done on the image of interest to make it suitable for further processing [6]. The convective image is subtracted from the reference image to extract only the convective portion of the image [5]. The gray level image is then converted to binary image using soft thresholding algorithm using the mean value of the image as the threshold value. The convective portion of the binary image will be white and nonconvective portion will be black. Morphological operations such as image opening and image closing depend on the ordering of the pixel values but not on the intensity of the pixel values. The size and

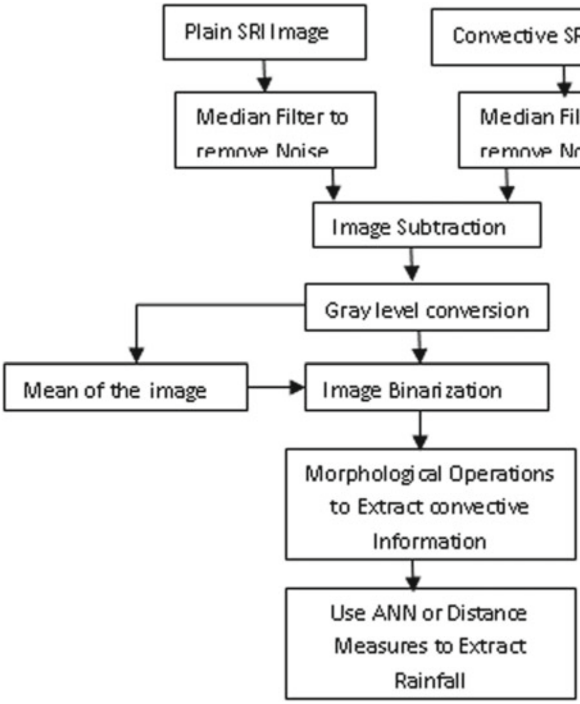


Fig. 2 Flow chart for the extraction of rainfall from the DWR SRI image

shape of the structuring element are sensitive to shapes in the input image. Image opening eliminates the small islands and sharp peaks. Image closing merges the narrow breaks and eliminates the holes. The morphological operated binary image is multiplied with noise eliminated original image which presents only convective activity of the image as shown in Fig. 3.

The extracted convective portion is compared against the color bar to extract the rainfall intensity over a particular location of the image [7]. The rainfall is extracted from the SRI image using various distance measures such as Euclidean, Standard Euclidean, City block, Minkowski, Chebychev, Mahalanobis, cosine, and correlation. This method presents an accuracy of 94%. Artificial Neural Networks are trained with color bar values to extract the rainfall intensity. Backpropagation algorithm is used to adjust the hidden layer and output layer weights. The ANN is trained with different backpropagation algorithms such as Levenberg–Marquardt, Gradient descent, Gradient Descent with Momentum, Conjugate gradient, Variable Learning Rate with Momentum, and resilient back propagation [8]. The Levenberg–Marquardt algorithm given the best output with minimum error and an accuracy of 98%.

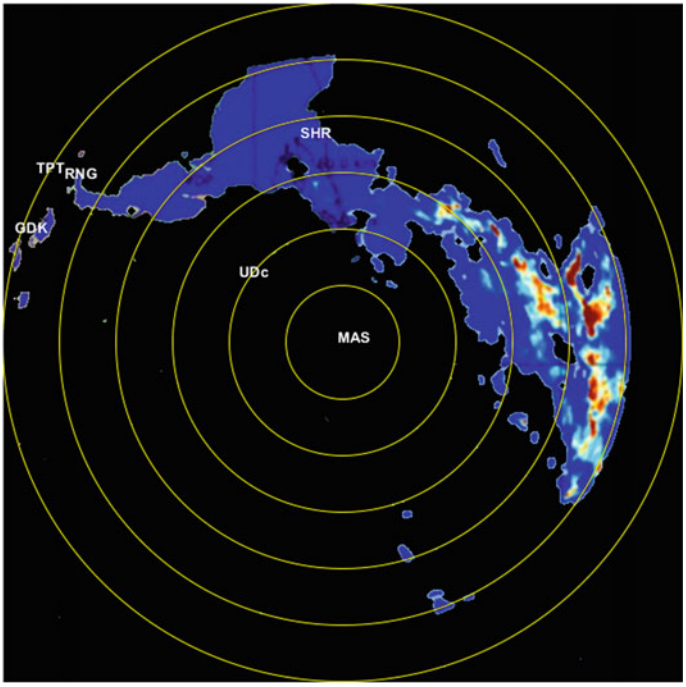


Fig. 3 Convective portion extracted from SRI image

4 Results and Discussions

The DWR SRI images were compared with the reference image to obtain the convective portion of the image by eliminating the static background is shown in Fig. 3. The green circles with increments of a 20 km radius are drawn with Chennai as the center. The static background image has no convective activity implies zero rainfall rate. The low-level clouds precipitate to a little extent. The deep convective clouds precipitate thunderstorm rain to a larger extent.

A Graphical User Interface (GUI) is Developed for Precipitation estimation from the DWR SRI product in MATLAB. It has a provision for rainfall estimation of a single image and a time series estimation of rainfall. GUI has the option to estimate the rainfall over a location (Latitude–Longitude) or click-based on the image. GUI fetches the location information from Google based on the lat-long entered in it. The Sample GUI for single image and time series extraction of rainfall is shown in Figs. 4 and 5. The images will be updated on the website for every 10 min. The images downloaded from the official website may not be continuous, and this is due to lack of updating the information. The lack of image or nonconvective activity over a location is treated as a NaN (Not a Number) for representation. The correlation coefficient is calculated without considering these NaN values.

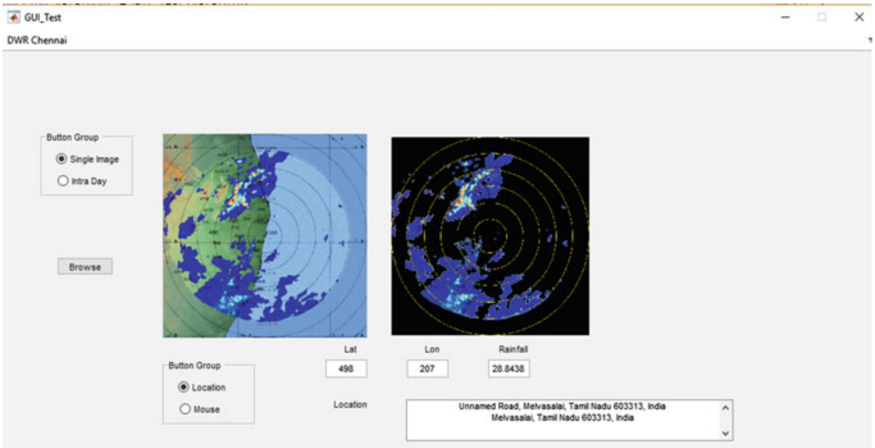


Fig. 4 GUI for a single image

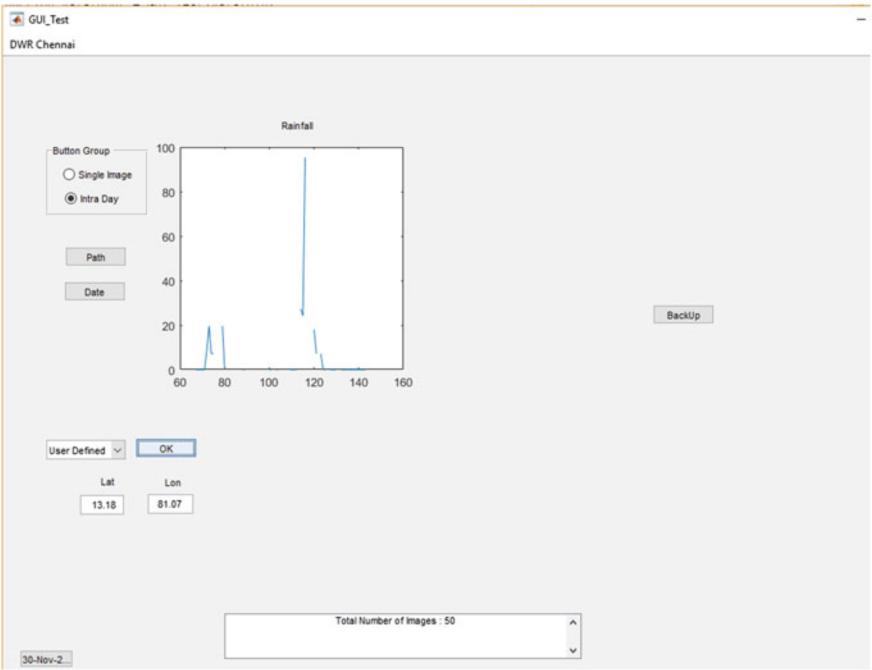


Fig. 5 GUI for a time series intraday

The rainfall from different AWS stations surrounding Chennai is collected and compared with the DWR estimated precipitation [2]. The metrics used in this paper are correlation coefficient. Correlation Coefficient determines the relationship

Table 1 Correlation coefficient between AWS and DWR

Date	Station	Correlation coefficient
3rd June 2016	SHAR	0.76515
	Chennai	0.69587
27th May 2016	SHAR	0.78128
	Chennai	0.73183
29th Nov 2015	SHAR	0.78424
	Chennai	0.79156
30th Nov 2015	SHAR	0.77818
	Chennai	0.76183

between the local average rainfall and rainfall of AWS and DWR derived rainfall is given in Eq. 2. It can be interpreted as the proportion of variance in y to the variance in x. The range of coefficient values is between -1 and 1. If the value of correlation coefficient is negative, it means as x increases y decreases in the same proportion. If the value of the correlation coefficient is positive, then x and y are positively correlated. If the value is zero, then there is no relation between input and output variables [9].

$$r = \frac{\sum(x - \bar{x})(y - \bar{y})}{\sqrt{\sum(x - \bar{x})^2 \sum(y - \bar{y})^2}} \tag{2}$$

The correlation coefficient between AWS and DWR extracted rainfall from SRI product are tabulated in Table 1 which shows the high degree of similarity. The correlation value is less due to the nonavailability of the DWR images at sometimes because of the problem in reception of the data due to glitches in the system and artifact effect arises due to temperature variation while receiving the images. The DWR station in Chennai has to update the images in the database for every 10 min, so a total of 144 images should be available per day. The station will not update the images in the web database. The nonavailability of images present wrong interpretation of the results. The nonavailability of the images will be represented as Not A Number (NAN), so there will be gaps in the time series image as shown in Fig. 5.

5 Conclusion

In this paper, precipitation is estimated from the DWR SRI images using image processing techniques. A GUI was developed for this application is user-friendly, which extracts the point rainfall or time series rainfall over a particular location. The extracted rainfall from the DWR was compared with the AWS data and a reasonable correlation coefficient was achieved. The method presented here is robust to noise.

Acknowledgements We are thankful to Mr. Kannan from IMD Chennai, Mr. Yesubabu from NARL Gadanki, and Mr. M. S. Arunachalam from IIT Chennai, who helped us in understanding the DWR images and deriving the algorithm.

References

1. Doviak RJ, Zrnik DS (1992) Doppler radar and weather observations, 2nd edn
2. Josephine VS, Mudgal BV, Thampi SB (2014) Applicability of Doppler weather radar based rainfall data for runoff estimation in Indian watersheds—A case study of Chennai basin. Sadhana 39(4):989–997
3. Suresh R, Ravichandran PK, Gupta JP, Thampi SB, Kalyanasundaram S, Rajesh Rao P (2005) On optimum rain rate estimation from a pulsed Doppler weather radar at Chennai. Mausam 56:433–446
4. Marshall JS, Palmer W (1948) The distribution of raindrops with size. Adv Geophys 2:1–56
5. Anil Kumar P, Anuradha B, Arunachalam MS (2017, July) Extraction of time series convective cloud profile from Doppler weather radar MAX (Z) product using a novel image processing technique. Int J Adv Eng Res Dev 4(7)
6. Gonzalez RC, Woods RE (1992) Digital image processing. Addison-Wesley, Reading, MA, p 716
7. Anil Kumar P, Anuradha B (2017) Estimating reflectivity of DWR images by analysing different colour spaces through distance measures. Adv Comput Sci Technol 10(8):2191–2200
8. Taravat A, Del Frate F, Cornaro C, Vergari S (2014) Neural networks and support vector machine algorithms for automatic cloud classification of whole-sky ground-based images. IEEE Geosci Remote Sens Lett 12(3):666–670
9. Kaur A, Kaur L, Gupta S (2012) Image recognition using coefficient of correlation and structural similarity index in uncontrolled environment. Int J Comput Appl 59(5)

Text-Independent Handwriting Classification Using Line and Texture-Based Features

T. Shreekanth, M. B. Punith Kumar and Akshay Krishnan

Abstract This paper addresses the problem of making a machine recognize the writer by means of the handwriting. It delineates the preprocessing methods used to enhance handwritings, so as to ease the process of feature extraction. It discusses six statistical texture-based features that characterize a handwriting. Once these features are extracted, a nearest neighbor approach is used to classify a sample handwriting into one of those in the database. The methods are verified on a self-compiled database, and a performance evaluation is also performed. This method can be used to identify an unknown handwriting and is in specific demand in the forensic domain. Unlike other biometric identification methods, handwriting-based identification is the least intrusive.

1 Introduction

The problem of classifying handwritings is one that has had a long history in research [1]. Earlier research focused on extracting data from handwritings [2–4]. In recent times, the problem of identifying the author has become just as significant as that of identifying data in the writing. In applications like forensics, it is often necessary to identify the writer based on the handwriting obtained from crime sites. Other applications include signature and document verification [5]. With applications that call for accuracy, it is very necessary that the solutions to such problems are reliable.

Handwriting recognition is characterized into two methods: text-dependent and text-independent. Text-dependent handwriting recognition involves the detection of

T. Shreekanth · A. Krishnan (✉)
Department of ECE, SJCE, Mysore, India
e-mail: akshay.krishnan.30@gmail.com

T. Shreekanth
e-mail: shreekanth_t@sjce.ac.in

M. B. Punith Kumar
Department of ECE, PESCE, Mandya, India
e-mail: punithpes@gmail.com

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_22

the characters used in the handwriting database before the writer can be recognized. This also requires the segmentation of characters as a preprocessing step. Once the characters have been segmented, they are classified. The differences between an individual's alphabet and the general alphabet are used to identify the handwriting.

In text-independent handwriting, the recognition is done without comparing the characters in the sample with those in the database. For this, texture-based statistical features are extracted from the dataset. These features are then used by a classifier to identify the sample image. The classifier may use nearest-neighbor approach or may be based on neural networks.

In this paper, we first present certain methods to enhance the quality of an image for handwriting recognition. This is followed by certain features that are extracted from the image in order to be fed to the classifier. The paper then discusses the nearest neighbors approach used for classification. This is followed by a discussion on the results obtained and future work.

2 Literature Survey

The preprocessing techniques used in handwriting recognition are similar to those of optical character recognition [6]. The difference is that in handwriting recognition the difference in handwriting (inter-class distance) is enhanced whereas in character recognition systems, the same is minimized. The preprocessing techniques used are as follows:

Converting to grayscale

Our implementation uses texture-based statistical features from images. Since the texture is independent of color, the classification and feature extraction is done on binary images. To convert a color image to a binary image, it is first converted to grayscale and is then thresholded.

Otsu's thresholding

Thresholding is used to convert the image to a binary form in which the foreground information is represented by black pixels and background information is represented by white pixels. Features can be easily extracted from binary images, once all the background information is isolated. The threshold that is selected plays a vital role is the thresholding process. In our implementation, Otsu's method [7] is used to select the threshold.

Connected component Labeling

Connected components labeling [8] is used to extract all the connected pixels of an image. Within a connected component, all pixels share the same intensity. The connected component labeling algorithm assigns a label (starting from 0) to each component. This step is necessary because it is much easier to apply feature extraction on selected connected components (letters or words) than on the image itself.

The algorithm works by scanning an image, pixel-by-pixel (both vertically and horizontally) in order to identify connected pixel regions V. (In our implementation, $V = \{255\}$).

Connectivity in this algorithm can be defined in terms of 8-neighbors or 4-neighbors. Our implementation uses 8-neighboured connected components. This is chosen owing to the sophisticated curves involved in handwritings.

Based on this information, the labeling of p occurs as follows:

- If all eight neighbors are 0, a new label is assigned to p.
- If one of the neighbors has an intensity value of 1, assign its label to the pixel.
- If more than one neighbors have an intensity value of 1, any value is assigned to p, and a note of this position is made.

Contour extraction using Moore neighborhood

The Moore neighborhood based tracing algorithm [9] that is used to extract contours is summarized as follows:

Input: A connected component X.

Output: A list Y consisting of the boundary pixels of each connected component.

Let Moore's neighborhood of pixel a in X be M(a).

The pixel being used, c is in M(p)

- Y is initialized to an empty set.
- The image is scanned until a black pixel in X is found.
- This pixel is added to Y.
- This point is set as the current point, say, s.
- Select the pixel from which s was entered.
- Let c be the next clockwise pixel.
- While c is not equal to s, loop:

```
If c is zero:
   Add c to Y
   Let p be c
   Move the current pixel c to the pixel from which p was entered.
Else:
   Move on to the next clockwise pixel.
```

Feature extraction: Research in handwriting recognition has brought to the forefront many features, most of them being statistical. Table 1 shows a summary of the features that have been used so far. Our solution uses statistical texture-based features that can uniquely identify a handwriting based on the research by [10].

Classification: The features corresponding to a sample handwriting can be classified as belonging to one from the database using the nearest neighbor approach or neural networks. Our implementation uses a nearest neighbor approach based on Euclidean distances.

Table 1 Literature survey of developments in handwriting recognition

Sl. No.	Paper name	Authors	Features used	Classifier
1	Writer identification using edge-based directional features [11]	Marius Bulacu Lambert Schomaker Louis Vuurpijl	Edge direction, edge hinge direction, run lengths, autocorrelation	Nearest neighbor, using Euclidean distances
2	Writer identification using text line based features, IEEE 2001 [12]	U. V. Marti, R. Messerli, H. Bunke	Relative heights of three main writing zones, slant angle of writing	K-Nearest neighbor, feed forward neural network
3	Biometric personal identification based on handwriting [13]	Yong Zhu, Tieniu Tan, Yunhong Wang	Gabor filters	Weighted euclidean distance
4	Individuality of handwriting [14]	Sargur N. Srihari, Sung-Hyuk Cha, Hina Arora, Sangjik Lee	Measures of pen pressure, measures of writing movement, measures of stroke formation, slant and proportion, paragraph and word level features	Nearest neighbor rule
5	Text-independent writer identification and verification using textural and allographic features [10]	Marius Bulacu, Lambert Schomaker	Direction and hinge probability distribution function, run length, autocorrelation, allograph level features	Nearest neighbor
6	Preprocessing and feature extraction for a handwriting recognition system [6]	T. Caesar, J. M. Gloger, E. Mandler	Peak area, upper line area, medium area, below baseline area, baseline area	N/A

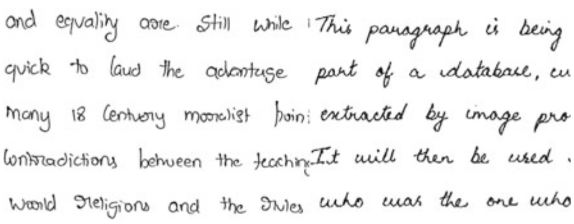


Fig. 1 Two handwritings from the database: before and after thresholding

3 Methodology

The process of text-independent handwriting recognition used in this paper can be divided into three subprocesses:

1. preprocessing of handwriting images.
2. Feature extraction.
3. Classification.

Preprocessing of handwriting images: Preprocessing is carried out as a series of spatial domain techniques listed as under:
Conversion to gray scale: Color images are converted to gray scale since color information is irrelevant to our approach and processing of gray scale images are much easier.
Thresholding: Features are easily extracted by separating the background information from the foreground. This is done by associating a 1 with the text and a 0 with the background, using Otsu's thresholding algorithm [7]. Our implementation later converts the text to a pixel value of 255 and the background to 0. This eases the processing and helps distinctly view the text. Figure 1 shows two handwritings from the database, one with thresholding and the other without it.
Connected components labeling: Since processing of all the text in an image of a paragraph is cumbersome, it is preferred to segment different portions of the paragraph and analyze them one after another. Depending on the handwriting, connected component labeling may segment words, or individual letters. A label is associated to each connected component at the end of this stage.
Contour extraction: In order to extract features like slant and roundness of the handwriting, contours need to be processed. This method is inspired by [9], where the features extracted are probability distribution functions from contours.
Feature extraction: The performance of a classification depends on the number and relevance of features extracted. The features extracted much be non-redundant. That is, each feature must convey information that is not conveyed by another. The features taken together must also provide enough information to distinguish one handwriting class from another. Taking into consideration various aspects of a handwriting, the following texture-based features are extracted.

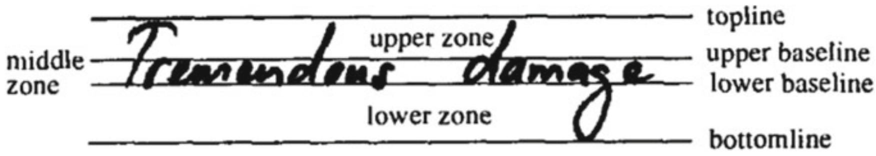


Fig. 2 Writing zones and bounding lines [12]

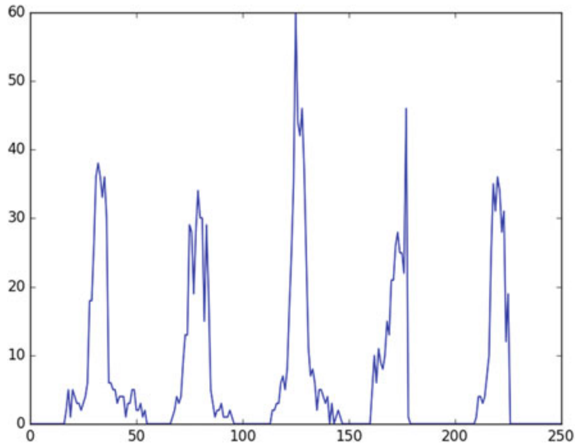


Fig. 3 The line profile/horizontal histogram of a handwriting with 5 lines of text, used for extracting the line-based features

1. **Three text line based features**

The features are mainly based on visible characteristics of the writing, such as the height of the three main writing zones, the width of the characters, the slant of the writing, and the text's legibility. These features can also be observed and interpreted by humans.

The features are based on the height of the three main writing zones, which are determined by the topline, the upper baseline, the lower baseline, and the bottomline (see Fig. 2). To determine these lines, a horizontal projection p of the text line image is computed. In the horizontal projection (as in Fig. 3), there is a spike between the upper baseline and the lower baseline, and relatively small values (distinguished by a threshold 't'. t is relative and varies from image to image. Therefore, it is set as a fraction of the peak value between the upper and lower baselines) between the lower baseline and bottomline, and the topline and upper baseline. Between the bottomline of one line and topline of the next, the value is zero.

$$x = |\text{topline} - \text{upper baseline}| \tag{1}$$

$$y = |\text{upper baseline} - \text{lower baseline}| \tag{2}$$

$$z = |\text{lower baseline} - \text{bottomline}| \tag{3}$$

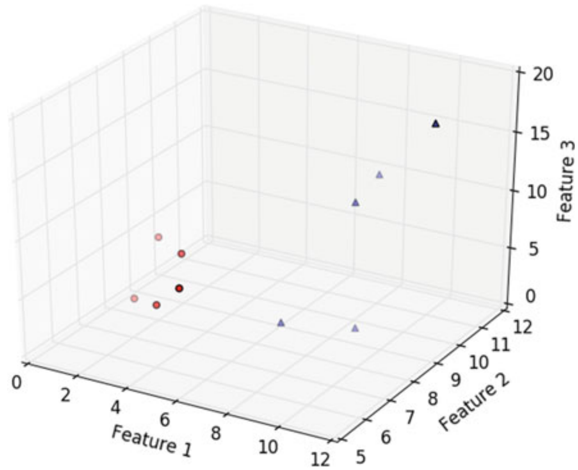


Fig. 4 Line-based features corresponding to two different handwritings

The values of x, y, and z however are scale variant. To achieve scale invariance, their relative values are considered as features. These are

$$f1 = x/y \tag{4}$$

$$f2 = y/z \tag{5}$$

$$f3 = x/z \tag{6}$$

The three features $f1$, $f2$, and $f3$ extracted for 8 different samples belonging to two handwritings (4 sample per handwriting), produced results as shown in Fig. 4.

2. **Contour-based probability distribution functions as features:**

Further, in this section, we describe the extraction methods for the texture-level features [10]. In these features, the handwriting is merely seen as a texture described by some probability distributions.

1. **Writing slant**: The slant of any handwriting is perhaps the most obvious feature that sets apart different writers. In order to capture the slant of a handwriting in a feature, we resort to contours.

In order to extract this feature, we follow these steps:

- For each contour in the image:
- For each point p1 along the contour:
- Select a point p2 that is a minimum number (k) away from p1 along the contour.
- K is decided based on the thickness of the handwriting and in our implementation, it is set to 4.
- Calculate the angle made by the line joining p1 and p2 with the horizontal.

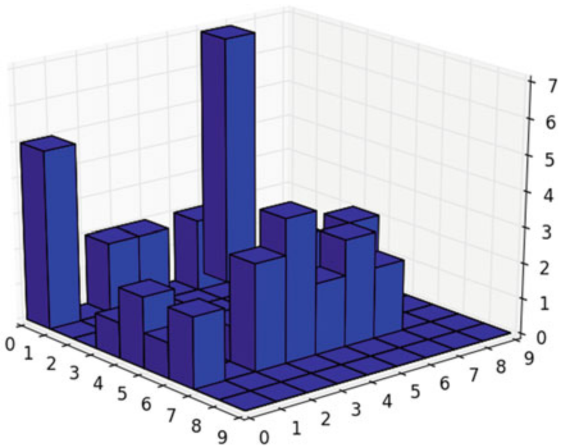


Fig. 5 Representation of the two-dimensional probability distribution function for one handwriting

- For computational ease, this angle is divided into groups of 20.

A number of contours corresponding to a particular angle are stored in a histogram which is normalized to give a probability distribution function, which is used as the fourth feature $f4$.

1. **Handwriting curvature**: This feature is used to identify the curvature of the handwriting. The following steps are to be followed for extracting this feature:

   - For each contour in the image:
   - For each point p1 along the contour:
   - Select a point p2 that is a minimum number (k) away from p1 along the contour.
   - Calculate the angle made by the line joining p1 and p2 with the horizontal, let it be A.
   - Select a point p3 that is a minimum number (k) away from p1 along the contour in the other direction.
   - Calculate the angle made by the line joining p1 and p2 with the horizontal, let it be B.
   - For computational ease, A and B are divided into groups of 20.
   - Increment the value for the particular group in the joint probability distribution function P(A, B).

This joint probability distribution (Fig. 5) is used as the fifth feature $f5$.

1. **Directional Co-Occurrence of handwriting:**

This feature is used to identify the roundness of the handwriting. The following steps are used to extract this feature:

- For each contour in the image.

Table 2 Performance evaluation of the handwriting recognition algorithm

Number of handwritings in database	100
Number of samples tested	30
Number of correct recognitions	26
Number of incorrect recognitions	4
Accuracy of the method	86.67%

- For each point p1 along the contour.
- Select a point p2 that is horizontally opposite to along the contour in the other direction. This is done by identifying all other points along the contour with same row-index and removing the ones that are redundant.
- Find the contour direction for both p1 and p2, let them be A and B.
- For computational ease, A and B are divided into groups of 20.
- Increment the value for the particular group in the joint probability distribution function $P(\mathrm{A}, \mathrm{B})$.
- This joint probability distribution constitutes the fifth feature $f5$.

Classification: The final stage of handwriting recognition is the classification of any handwriting sample into a class that is defined by the handwritings in the database. Our implementation uses a classifier that minimizes the variance between the sample and the classes in the database, in terms of the six features identified in the previous step.

4 Results

The database used for handwriting recognition in our implementation comprised 100 handwritings of which 6 were self-compiled and the others were from the IAM handwriting database [15]. Once the algorithm was implemented, it was tested using different handwriting samples of the same authors. The results are summarized in Table 2.

5 Conclusion and Future Work

In order to increase the accuracy of the handwriting recognition system, more features can be incorporated into the system. Psychology has suggested characteristics of a person's handwriting based on the pressure applied on the paper [14]. These manifest in the form of the thickness of the handwriting. This paper considers only those features based on the overall texture of the handwriting. If the thickness were to be evaluated, the information obtained can be used to further improve the performance of the classifier.

The paper focuses on off-line handwriting recognition in which the handwriting can be recognized only if a written paragraph is fed as a sample. An online recognition system would be able to detect the writer, during the process of writing itself. This can be used in tablets and other such touchscreen devices. Future work shall focus on extending our methods to on-line handwriting recognition.

References

1. Raj A, Chaudhary A (2016) A review of personal identification using handwriting. Int J Eng Sci Comput
2. Feature extraction and identification of handwritten characters, Computer Technology Application Key Laboratory of Yunnan Province, Kunming University of Science & Technology, Kunming, China
3. Nath RK, Rastogi M (2012) Improving various off-line techniques used for handwritten character recognition: a review. IJCA 49(18)
4. Saady YE, Rachidi A, El Yassa M, Mammass D (2011) Amazigh handwritten character recognition based on horizontal and vertical centerline of character. IJAST 33(17):33–50
5. Nguyen V, Blumenstein M (2011) An application of the 2D gaussian filter for enhancing feature extraction in off-line signature verification. In: 2011 international conference on document analysis and recognition (ICDAR), pp 339–343
6. Preprocessing and Feature Extraction for a Handwriting Recognition System. 0-8186-4960-7/93 Q 1993 IEEE
7. Kurita T, Otsu N, Abdelmelek N (1992) Maximum likelihood thresholding based on population mixture models. Pattern Recognit 25(10):1231–1240
8. He L, Chao Y, Suzuki K, Wu K (2009) Fast connected-component labeling. Pattern Recognit (Elsevier)
9. Seo J, Chae S, Shim J, Kim D, Cheong C, Han T-D. Fast contour-tracing algorithm based on a pixel-following method for image sensors
10. Bulacu M, Schomaker L (2007) Text-independent writer identification and verification using textural and allographic features. IEEE Trans Pattern Anal Mach Intell 29(4)
11. Bulacu M, Schomaker L, Vuurpijl L (2003) Writer identification using edge-based directional features. In: Proceedings of 7th international conference on document analysis and recognition (ICDAR 2003), IEEE Press
12. Marti U-V, Messerli R, Bunke H (2001) Writer identification using text line based features. Institut fur Informatik und angewandte Mathematik, 0-7695-1263-1/01 0, IEEE
13. Zhu Y, Tan T, Wang Y. Biometric personal identification based on handwriting (unpublished)
14. Srihari SN, Cha S-H, Arora H, Lee S (2002) Individuality of handwriting. J Forensic Sci 47(4)
15. Marti U, Bunk H (2002) The IAM-database: an english sentence database for off-line handwriting recognition. Int J Doc Anal Recognit 4:39–46

A Unified Preprocessing Technique for Enhancement of Degraded Document Images

N. Shobha Rani, A. Sajan Jain and H. R. Kiran

Abstract The field of Document Image Processing has encountered sensational development and progressively across the board relevance lately. Luckily, propels in PC innovation have kept pace with the fast development in the volume of picture information in different applications. One such utilization of Document picture preparing is OCR (Optical Character Recognition). Pre-preparing is one of the pre-imperative stages in the handling of record pictures which changes the archive to a frame reasonable for ensuing stages. In this paper, various preprocessing techniques are proposed for the enhancement of degraded document images. The algorithms implemented are adept at handling variety of noises that include foxing effect, illumination correction, show through effect, stain marks, and pen and other scratch marks removal. The techniques devised works based on noise degradation models generated from the attributes of noisy pixels which are commonly found in degraded or ancient document images. Further, these noise models are employed for the detection of noisy regions in the image to undergo the enhancement process. The enhancement procedures employed include the local normalization, convolution using central measures like mean and standard deviation, and Sauvola's adaptive binarization technique. The outcomes of the preprocessing procedure is very promising and are adaptable to various degraded document scenarios.

N. Shobha Rani (✉) · A. Sajan Jain · H. R. Kiran
Department of Computer Science, Amrita School of Arts and Sciences,
Amrita Vishwa Vidyapeetham, Mysuru, India
e-mail: n_shobharani@asas.mysore.amrita.edu

A. Sajan Jain
e-mail: jain.sajan2@gmail.com

H. R. Kiran
e-mail: kiranhr1993@gmail.com

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_23

1 Introduction

Records that can be handled by OCR incorporate antiquated, matured archives, for example, chronicled books, machine printed archives, for example, reminders, letter specialized reports, and books, hand composed archives, for example, individual letters, addresses on postal mail, notes in the edges of records, online written by hand archives, and so on.

Among all the document types, ancient documents of machine printed type or handwritten types are highly affected with noise due to aging [1] and hence in need of intensive preprocessing. The degradations incur due to noise are categorized as pen marks, stain marks, scratch marks, scan motion blur, bleed-throughs, nonuniform illumination, show through effects and foxing effect, etc. [2, 3].

Preprocessing of document image varies based on the document context leading to the design of algorithms suiting the document contexts. Quite a good number of works are reported in this area including text strokes crossing removal [3], graphical component detection and removal [4], adaptive document binarization [5], preprocessing for handwritten documents [6], page layout analysis [7], neural network based document preprocessing [8], stain removal [9], enhancement of degraded document quality [10], and many other methods. The details of some of the important works are as summarized subsequently.

Chang et al. [11] had proposed an approach for picture denoising and pressure which utilizes a versatile, information-driven limit by means of wavelet soft thresholding. The limit was inferred in a Bayesian structure, and the earlier utilized on the wavelet coefficients was the summed up Gaussian conveyance. The experiments depicts that BayesShrink, was regularly within 5% of the Mean Squared Deviation of the best delicate thresholding benchmark with the picture expected. Hsia et al. [12] had propped a strategy to denoise the interfering or obvious watermarks from foundation data for written by hand archives utilizing versatile histogram adjustment to stifle undesirable meddling strokes and acquired an unmistakable picture utilizing versatile directional lifting-based discrete wavelet changes. The trial comes about portray that Haar wavelet was not the ideal technique in light of the fact that the first strokes were broken, however, the meddling strokes were expelled. Ntogas et al. [13] had proposed a binarization method for pre-filtered historical manuscripts images consisting of five discrete steps from data acquisition, conversion of RAW files to JPEG, cropping and converting to gray scale and then denoising using various filters in spatial domains, followed by thresholding and morphology. Experimentations were performed on Byzantine historical manuscripts. Kitadai et al. [14] had proposed a similarity evaluation technique for character designs with missing shape parcels utilizing nonlinear normalization and adjusted the formats for every trial of the recovery. The result of the strategy indicated changes of the recovery precision up to 72%. Shirai et al. [15] had concocted a strategy for anisotropic morphological enlargement by means of certain smoothing to restore the corrupted character states of binarized pictures. They connected a smoothing technique not to the double picture but rather to the separation changed picture, and after that reconverted it by binarization. Additionally, depicted a strategy for reestablishing the flood of strokes, utilizing verifiable smoothing for the morphological task in order to evacuate loud examples out of sight and interface the fractionated character strokes. Lu et al. [16] had contributed a binarization procedure for sectioning content from seriously lit up archive pictures by using smoothing polynomial surface for the shading estimation and pay, which delivered a generally consistently lit up record picture that could be binarized by some worldwide thresholding strategies. Experimentation demonstrated that the system was tolerant to the varieties in content size and report differentiate could binarize severely lit up archive picture. Kavallieratou and Antonopoulou [17] had built up a recursive calculation for improvement of recorded archives by utilizing mixes of entangled picture handling strategies by considering spatial qualities of the report pictures. Normal issues of verifiable reports like foundation difference stains and straightforwardness were overwhelmed by their technique in correlation with Bernsen's, Niblack's, and Otsu's strategies for binarization and record improvement. Wolf [18] had proposed another technique for blind document bleed-through removal attributed of three phases: making of a twofold MRF display with a solitary perception field, at that point plan of an iterative streamlining calculation in light of the base cut/most extreme stream in an entire rebuilding process. The technique is assessed on checked record pictures of the eighteenth century and 83.23% review and 74.85% accuracy rates were asserted. Shi and Govindaraju [19] had proposed a standardization calculation appropriate for recorded report pictures utilizing a versatile direct capacity to estimate the uneven foundation because of the uneven surface of the archive paper, matured shading and light wellspring of the cameras for picture lifting. The calculation adaptively caught the foundation with a "best fit" straight capacity and standardized regarding the estimation. Garain et al. [20] had contrived a system that isolates the closer view and foundation in low-quality record pictures experiencing different sorts of debasements including examining clamor, maturing impacts, uneven foundation, or frontal area, and so forth. The calculation is likewise versatile to handle those issues of uneven enlightenment and nearby changes or non-consistency in foundation and frontal area hues. The approach was basically intended for preparing of shading archives and assessment results demonstrate that the strategy could remove lines and words with correctness's of around 84 and 93%, individually. Most promising works are reported in the literature suiting the requirements of a wide variety of documents of varying layouts and contexts. The algorithms devised are mostly adaptable with gray scale/monochrome image based operations. It is noticed that commonly found problems of degradation types addressed include illumination correction, bleed-through removal, aging effects removal, and other binarization-related issues. Preprocessing techniques devised are based on adaptive thresholding, morphological operations based and normalization techniques. In the perspective of improvising the preprocessing outcomes, a novel technique works based on the knowledge of noise degradation models of various degradation types persisting in the noisy document images. The working of proposed techniques is discussed subsequently.

2 Proposed Technique

The techniques for preprocessing of degraded document types include the creation of noise degradation models, detection of foxing effects (noisy regions) in the images, nonuniform illumination correction, show through effect removal, pen, scratch, and other stain mark removals.

2.1 Creation of Noise Degradation Models

Image enhancement is a nontrivial process which is highly subjective and bound by the context of image contents. Image restoration is the process of remodeling the actual geometrical properties of an image based on the prior knowledge of noisy phenomenon of a degraded image. Creation of noise degradation models based on the noisy patterns in an image is one the image restoration techniques [21]. In the proposed method, Gaussian and Rayleigh noise degradation models along with image averaging and harmonic mean image restoration filters are utilized for the creation of noise degradation models. Figure 1 depicts some of the noise patterns administered for the creation of noise degradation models.

Figures 2, 3, and 4 represent the sample documents considered for experimentation.

The gray levels of the noisy patterns of the image are exploited for estimation of its statistical features. Let v represent the Gaussian random variable and also a gray level, μ indicates the mean of average value of v and σ is the standard deviation, and then the Gaussian noise degradation estimate is given by (1).

$$p(v) = -\frac{1}{\sqrt{2\pi\sigma}} e^{(v-\mu)^2} / 2\sigma^2 \tag{1}$$

Fig. 1 Instances of noise patterns

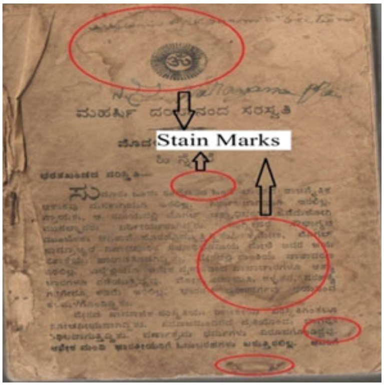


Fig. 2 Degraded document image—type 1

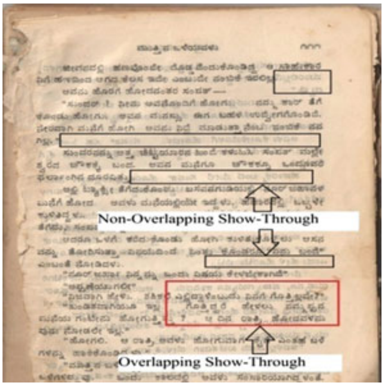


Fig. 3 Degraded document image—type 2

Similarly, the Rayleigh noise degradation estimate with respect to a random variable v is given by (2).

$$p(v) = \begin{cases} \frac{2}{b}(v - a)e^{-(v-a)^2/b}; v \geq a \\ 0; v < a \end{cases} \tag{2}$$

Further, the harmonic mean filter a image averaging is employed as one of the prominent features for creation of feature vector at the later level. If $g(x, y)$ is the pixel location for which the average gray level with respect to a region of K pixels, then average gray level at location $g(x, y)$ is given by (3).

$$\overline{g}(x, y) = \frac{1}{K}\sum_{i=1}^{K} g_i(x, y) \tag{3}$$

where $\overline{g}(x, y)$ is the average gray level computed over a neighborhood of K pixels with respect to a pixel $g(x, y)$ at location i.

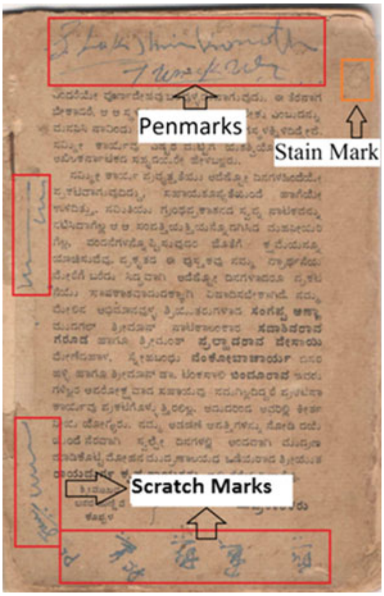


Fig. 4 Degraded document image—type 3

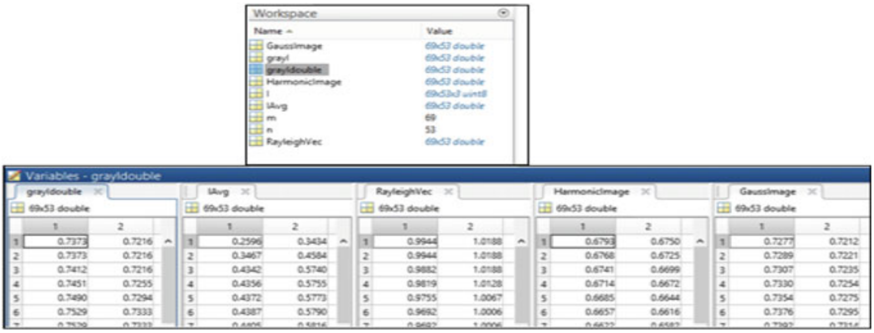


Fig. 5 Features extracted from noisy patterns

The harmonic mean filtering operation is given by (4)

$$\overline{g}(x, y) = \frac{mn}{\sum_{(s,t)\in S_{xy}} \frac{1}{g(s,t)}} \tag{4}$$

where mn represents the size of rectangular neighborhood S_{xy}, $g(s, t)$ is the pixel at location (s, t) and $\overline{g}(x, y)$ is the harmonic mean at location $g(s, t)$.

Figure 5 presents the features extracted from the noisy patterns of the images.

2.2 Detection of Foxing Effect

The features extracted corresponding to noise degradation models and mean filters are further subject to detection of noisy patterns through convolution technique [22]. The features extracted are convolved over the image for the detection of foxing effects in the image. If the features computed during convolution matches with feature vectors of degradation models, then the regions will be considered as noisy.

Figure 6 represents the detection of foxing effects for an instance of an image through convolution and Fig. 7 shows the other types of noisy patterns detected using convolution.

Convolution is carried out with rectangular neighborhood of dimensions 15×15. Once the noisy regions are detected, the removal of detected noise is accomplished by comparing the mean and standard deviation range of noisy pixels.

Let μ and σ represent the mean and standard deviation, then the probability that an arbitrary pixel $p(x, y)$ to get assigned with the value between 0 and 1 is given by (5).

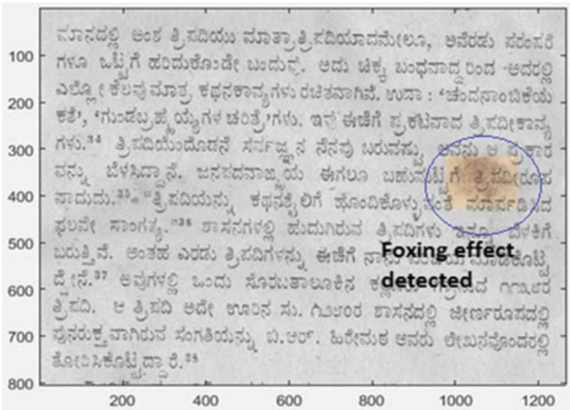


Fig. 6 The noisy region detected is encircled

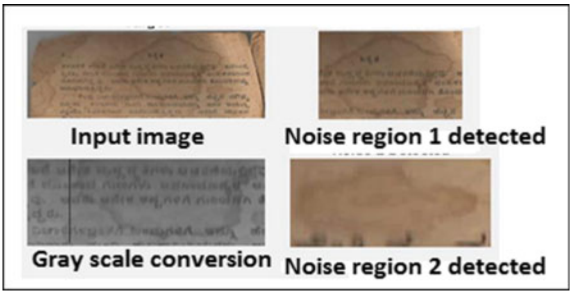


Fig. 7 Samples of detected noise patterns

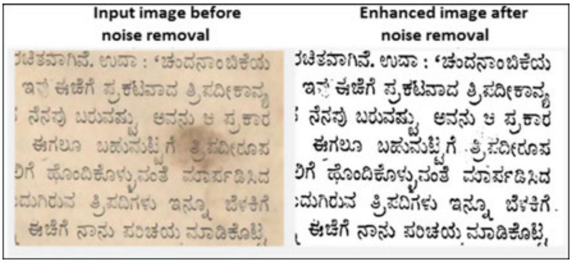


Fig. 8 Removal of foxing effect

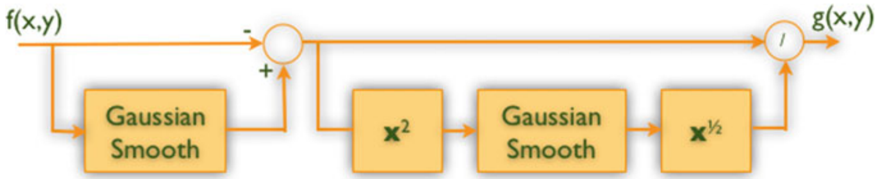


Fig. 9 Procedure of local spatial smoothing

$$p(x, y) = \begin{cases} 1 \\ 0 \end{cases} \tag{5}$$

where 1 indicates the background and 0 indicates the text pixels. It is expected $p(x, y) = 1$ subject to $0.3 < \mu[p(x, y)] < 1$ otherwise $p(x, y) = 0$.

Figure 8 depicts the outcome of foxing noise removal.

2.3 Nonuniform Illumination Correction

The occurrences of illumination inconsistencies within the document are one of the factors that results in noise coverage of textual regions. In the proposed method, correction of illumination is achieved using local spatial smoothing technique [23]. Local spatial smoothing is one of the most commonly used preprocessing techniques for medical image enhancement. Local spatial smoothing produces the effect of low pass filtering by attenuation of high-frequency components in an image. As in the current scenario, the illumination nonuniformities are existent due to the overriding of low-frequency regions (background) by high-frequency elements, therefore it is attempted to perform illumination correction using local spatial smoothing technique. Gaussian filtering [24] is the core functionality of spatial smoothing which is as presented in Fig. 9.

If $f(x, y)$ represents the input image, $m_f(x, y)$ and $\sigma_f(x, y)$ are the local mean and local standard deviation estimated on $f(x, y)$, then the output $g(x, y)$ is given by (6).

Fig. 10 Results of illumination correction

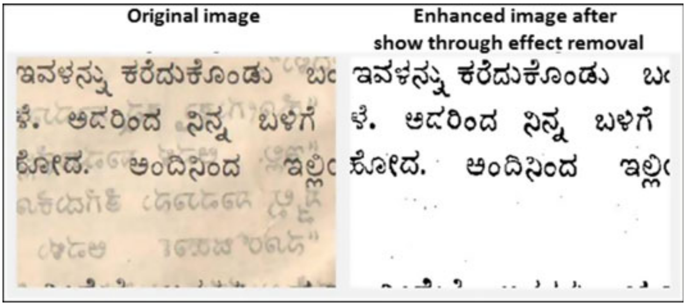


Fig. 11 Removal of show through effect

$$g(x, y) = \frac{f(x, y) - m_f(x, y)}{\sigma_f(x, y)} \tag{6}$$

Figure 10 depicts the outcome of illumination correction.

2.4 Removal of Show Through Effect, Pen Marks and Others

Elimination of show through effect, pen/scratch marks and stains due to aging in document images are also performed using local normalization through local spatial smoothing technique. Figures 11, 12, and 13 show the results of local normalization.

3 Experimental Analysis

It is well known, the evaluation of preprocessing algorithmic outcomes is subjective to the satisfaction of user, hence the performances of proposed techniques are eval-

Fig. 12 Removal of stain marks

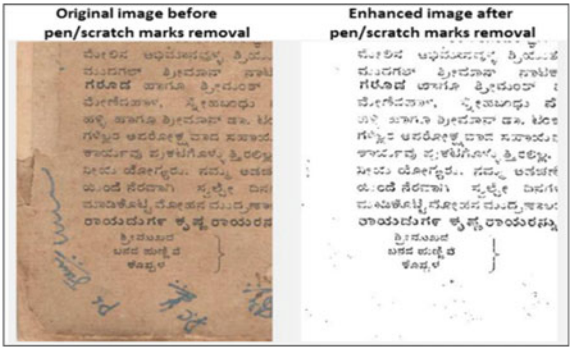


Fig. 13 Removal of pen/scratch marks

Table 1 Performance of preprocessing techniques

Type of noise pattern	No. of documents experimented	No. of documents with rating "Good"	No. of documents with rating "Moderate"	No. of documents with rating "Poor"
Foxing effect	72	33	28	11
Illumination correction	85	40	32	13
Show through effect	75	30	35	10
Pen/scratch marks	90	56	26	8
Stain removal	150	112	35	3

uated through subject studies. Around 50 scorers are used to rate the performance of outcomes obtained. All the datasets employed for experimentation are collected from Kannada library, where the context of works is related to poetry and literature of eighteenth and nineteenth centuries. A collection of more than 500 documents are scan converted to document images with a resolution of 300 dpi. The performance of outcomes is rated qualitatively into good, moderate and poor classes. The performance metrics of ratings provided by scorers are tabulated in Table 1 (Figs. 14 and 15).

Fig. 14 Accuracy of foxing effect removal

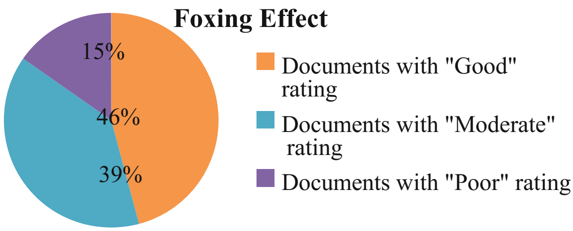


Fig. 15 Accuracy of illumination correction

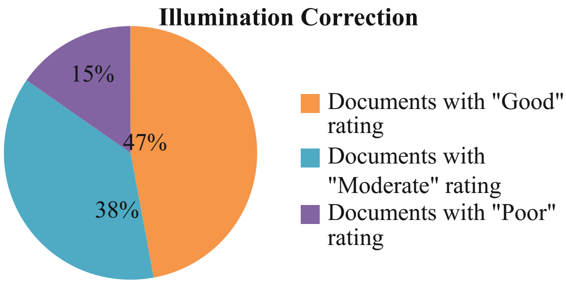
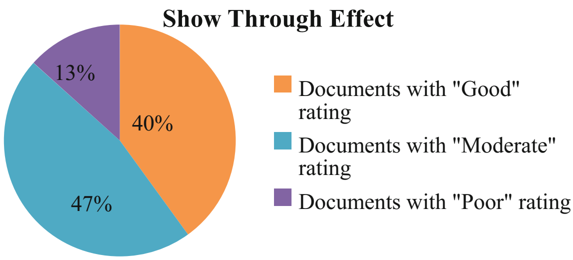


Fig. 16 Accuracy of show through effect



3.1 Conclusion

A total of 472 documents with various noise patterns were experimented and the results showed a good amount of accuracy in preprocessing all different kinds of noises. Stain mark removal showed the highest accuracy with 75% of the testing documents being rated as good. Followed by, pen/scratch mark removal, illumination correction, foxing effect removal and show through effect removal with 62, 47, 46, and 40% of the documents being rated good, respectively (Figs. 16 and 17).

4 Conclusion

The preprocessing of historical documents is different when compared with other type of images like simple text documents of machine printed or typewritten images, etc. In this research, the focus is on developing some generic preprocessing algorithms

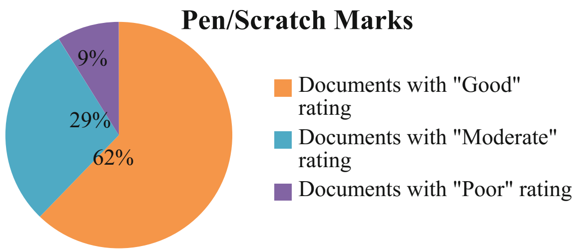


Fig. 17 Accuracy of pen/scratch marks

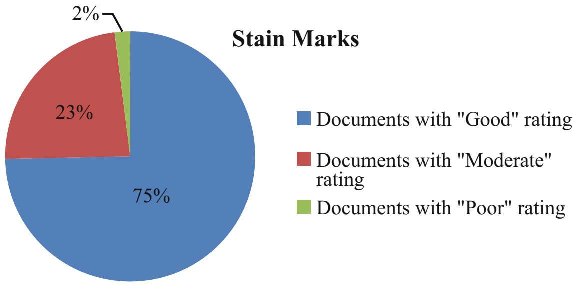


Fig. 18 Accuracy of stain marks removal

specific for ancient document images. Specifically, the problems include the elimination of noises, identification and removal of stains, scratch marks, pen marks or irrelevant portions of text which are not in understandable format neither by humans nor by machines. The proposed technique employs noise models like Gaussian, Lucy Richardson, Rayleigh, etc., and extracts features from the noise imaged followed by the training of the degraded images and detection using convolution and correlation. The experimentation was carried out on the South Indian language historical document images. The results were shown in the form of screenshots which showed clear visuals of noise being eliminated from the degraded Kannada historical documents (Fig. 18).

References

1. Gupta MR, Jacobson NP, Garcia EK (2007) OCR binarization and image pre-processing for searching historical documents. Pattern Recogn 40(2):389–397
2. Farooq F, Govindaraju V, Perrone M (2005) Pre-processing methods for handwritten Arabic documents. In: Proceedings of the eighth international conference on document analysis and recognition, 2005, pp 267–271. IEEE
3. Rani NS, Vasudev T (2018) An efficient technique for detection and removal of lines with text stroke crossings in document images. In: Proceedings of international conference on cognition and recognition, pp 83–97. Springer, Singapore
4. Rani DANS, Vineeth P, Ajith D (2016) Detection and removal of graphical components in pre-printed documents. Int J Appl Eng Res 11(7):4849–4856
5. Gatos B, Pratikakis I, Perantonis SJ (2006) Adaptive degraded document image binarization. Pattern Recogn 39(3):317–327

6. Farooq F, Govindaraju V, Perrone M (2005) Pre-processing methods for handwritten Arabic documents. In: Proceedings eighth international conference on document analysis and recognition, 2005, pp 267–271. IEEE
7. O'Gorman L (1993) The document spectrum for page layout analysis. IEEE Trans Pattern Anal Mach Intell 15(11):1162–1173
8. Rehman A, Saba T (2014) Neural networks for document image preprocessing: state of the art. Artif Intell Rev 42(2):253–273
9. Gatos B, Ntirogiannis K, Pratikakis I (2009) ICDAR 2009 document image binarization contest (DIBCO 2009). In: 10th international conference on document analysis and recognition, 2009. ICDAR'09, pp 1375–1382. IEEE
10. Kavallieratou E, Stamatatos E (2006) Improving the quality of degraded document images. In: Second international conference on document image analysis for libraries, 2006. DIAL'06, 10-pp. IEEE
11. Chang SG, Yu B, Vetterli M (2000) Adaptive wavelet thresholding for image denoising and compression. IEEE Trans Image Process 9(9):1532–1546
12. Hsia CH, Hoang HG, Tu HY (2015) Document image enhancement using adaptive directional lifting-based wavelet transform. In: 2015 IEEE international conference on consumer electronics-Taiwan (ICCE-TW), pp 432–433. IEEE
13. Ntogas N, Veintzas D (2008) A binarization algorithm for historical manuscripts. In: WSEAS Proceedings of the international conference on mathematics and computers in science and engineering, no. 12. World Scientific and Engineering Academy and Society
14. Kitadai A, Nakagawa M, Baba H, Watanabe A (2012) Similarity evaluation and shape feature extraction for character pattern retrieval to support reading historical documents. In: 2012 10th IAPR international workshop on document analysis systems (DAS), pp 359–363. IEEE
15. Shirai K, Endo Y, Kitadai A, Inoue S, Kurushima N, Baba H et al (2013) Character shape restoration of binarized historical documents by smoothing via geodesic morphology. In: 2013 12th international conference on document analysis and recognition (ICDAR), pp 1285–1289. IEEE
16. Lu SJ, Tan CL (2007) Binarization of badly illuminated document images through shading estimation and compensation. In: Ninth international conference on document analysis and recognition, 2007. ICDAR 2007, vol 1, pp 312–316. IEEE
17. Kavallieratou E, Antonopoulou H (2005) Cleaning and enhancing historical document images. In: International conference on advanced concepts for intelligent vision systems. Springer, Berlin, pp 681–688
18. Wolf C (2010) Document ink bleed-through removal with two hidden markov random fields and a single observation field. IEEE Trans Pattern Anal Mach Intell 32(3):431–447
19. Shi Z, Govindaraju V (2004) Historical document image enhancement using background light intensity normalization. In: Proceedings of the 17th international conference on pattern recognition. ICPR 2004, vol 1, pp 473–476. IEEE
20. Garain U, Paquet T, Heutte L (2006) On foreground—background separation in low quality document images. IJDAR 8(1):47
21. Kanungo T, Haralick RM, Phillips I (1993) Global and local document degradation models. In: Proceedings of the second international conference on document analysis and recognition, 1993. IEEE, pp 730–734
22. Lee J-S (1980) Digital image enhancement and noise filtering by use of local statistics. IEEE Trans Pattern Anal Mach Intell 2:165–168
23. Ord JK, Getis A (1995) Local spatial autocorrelation statistics: distributional issues and an application. Geogr Anal 27(4):286–306
24. Young IT, Van Vliet LJ (1995) Recursive implementation of the Gaussian filter. Sig Process 44(2):139–151

An Efficient Classifier for P300 in Brain–Computer Interface Based on Scalar Products

Monica Fira and Liviu Goras

Abstract In this paper, a simple but efficient method for detection of P300 waveform in a Brain–Computer Interface (BCI) is presented. The proposed method is based on computing scalar products between the waveforms to be classified and a P300 pattern. Depending on the degree of concentration of the subject and the number of trails, rates of recognition between 85 and 100% have been obtained.

1 Introduction

The Brain-Computer Interface (BCI) is a communication system where the messages or commands that a person sends to the outside world do not go through the normal pathways of the brain and peripheral muscle. In fact, in a BCI system built to function based on EEG signals, messages are encoded in the EEG activity. The purpose of a BCI system is to provide its user with an alternative way of acting on the world by extracting the encoded activity from the EEG signal.

A BCI system is built from the following blocks or component parts: (a) a set of sensors that record the neuronal activity; (b) a signal processing block that extracts features; and (c) a translation algorithm that creates commands to operate an external device. Another aspect that needs to be considered is that the loop needs to be filled with feedback from the external device to the BCI user. Therefore, a BCI system is a closed loop running in real time. In the case of BCI communications systems, the external device serves as a means of communicating the user with the external environment.

A large number of BCI systems have been reported in the literature over theyears. Their operating principles use different neuronal characteristics, so they differ as a way of using neuronal information collected from sensors. Thus, noninvasive EEG

M. Fira (✉)
Institute of Computer Science, Romanian Academy, Iasi, Romania
e-mail: mfira@etti.tuiasi.ro

L. Goras
"Gheorghe Asachi" Technical University of Iasi, Iasi, Romania

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_24

signals are use slow cortical potentials [1], motor potentials [2], event-related synchronizations and desynchronizations [3, 4], steady-state evoked potentials [5], and P300 potentials [6].

Depending on the way of communication between the human brain and the system, there are two classes of BCI systems, namely: dependent and independent. Dependent BCIs use the activity in the brain's normal output pathways to generate the brain activity (e.g., EEG) required for the system to function. An independent BCI does not depend on the brain's pathways (i.e., peripheral nerves or muscles); activity in such brain output is unnecessary to create the brain activity (e.g., EEG) required to execute a certain command [6, 7]. The independent BCIs are of greater theoretical interest than dependent BCIs [8–11]. This increased interest in independent BCI systems is due to the fact that they provide the brain with completely new output pathways comparative with dependent BCIs. Besides, for people with the most severe neuromuscular disabilities, who may lack all normal output channels (including extraocular muscle control), independent BCIs are likely to be more useful [12, 13].

Until now, P300-based BCI systems have been extensively studied and it has been concluded that from a psychological point of view, P300 evoked potentials that help maintain active memory when external stimuli are upgraded [7]. It has been shown that the P300 potential amplitude is directly proportional to the attention given to fulfilling a particular task [14, 15], and is associated to memory performance. The amplitude of the P300 pulse can also be seen as a reflection of the extent to which the information received from the outside is brain-processed in the form of representations of stimuli that generated the impulse. Amplitude variations are a measure of the degree to which information is processed. In short, there is a clear correlation between the impulse amplitude and the quality of the information processing. The pulse duration is the one that measures the classification speed, [16]. Tests have shown that the duration of the impulse varies from one subject to another and is a measure of the rapidity with which subjects allocate and maintain attention resources. There are pathologies in which P300 pulse duration is elevated, [15], and this is still an argument that it can measure cognitive abilities.

BCI based on the P300 can be done not only by illuminating letters or figures but also pictures or icons. The physiological mechanism of generating the P300 evoked potential is the same as in the case of the spelling paradigm, a visual stimulus. Thus, Hoffmann et al. [17] present a paradigm with the illumination of six pictures and on their web page, they also provide the EEG signals collected on 8 human subjects, with and without disabilities. They test two classification methods, namely, BLDA and FLDA as classifiers for P300 waveform detection.

In this paper, we present a simple and efficient method of classifying P300 evoked potentials by illuminating some pictures. To test the algorithm, we used the database available on the Internet at [18].

2 Materials and Methods

2.1 Experimental Schedule, Subjects, and Database

Experimental setup and schedule. Human subjects were shown six images on a laptop monitor. These images were: a television set, a telephone, a lamp, a door, a window, and a radio. The images were illuminated randomly, with only one image at a time. Image has been presented to the subject for 100 ms, and no more images have been lit in the next 300 ms, i.e., the inter-stimulus interval was 400 ms. The EEG signal was collected using the international standard 10–20 at 2048 Hz sampling rate and 32 electrodes placed [17].

For each subject, there were four sessions, two sessions a day and the time between the first session and the last session was less than two weeks. For each session, there were six rounds, one round for each image.

The subjects silently counted how often a picture appeared a prescribed image. After each run, the subjects were asked the result of counting. This has been done to monitor the performance of subjects [17].

Subjects. Human subjects were both normal and disabled, i.e., subjects 1–5 were people with handicaps and subjects 6–9 were normal. The disabled subjects were all wheelchair-bound but had varying communication and limb muscle control abilities [17].

Database. Four seconds after the alert tone, a random sequence of blinkers was started and the EEG registered. The flash sequence was randomized in block mode, which means that after six flashes, each image lit once, and after 12 flashes, each picture was blown twice, etc. The number of blocks was randomly chosen between 20 and 25. On average, 22.5 blocks of six intermittent were displayed in a single step, i.e., a cycle consisted of 22.5 target studies (P300) and $22.5 * 5 = 112.5$ nontarget (non-P300). The duration of a cycle was about one minute, and the duration of a session (including the installation of electrodes and short breaks between rounds) was about 30 min. A session covered, on average, 810 trials, and all data for one subject consisted, on average, of 3240 trials [17].

2.2 The Method

The first step consists of data preparation stage, namely the transformation of the EEG signals into sets of signals to be further processed. The second stage consists in choosing the EEG channels on which the classification will be made. This is followed by the step of *building a pattern* of the P300 waveform based on the EEG signals on the channels that were selected in the previous step. The next step is to classify the signals based on P300 pattern by calculating the *scalar product* between the pattern and each new EEG signal that is to be classified, and then establishing that the P300 has the signal corresponding to the illumination of the picture having the highest scaled product. In other words, at this stage, our classification problem appears as a 1 out of 6 ranking: which one of the 6 pictures has the highest probability to be the desired picture. For this, we will calculate the scalar product between the P300 pattern and the EEG signal corresponding to each photo illumination, and the decision was based on the maximum of the 6 scalar product values. In the following, we will detail each stage separately.

Preprocessing stage: For preprocessing EEG data, we chose to keep the same way of filtering, segmenting, windsorizing, and normalizing as Hoffmann used in [17]. Thus for data *filtering,* a 6th order forward–backward Butterworth bandpass filter was used. Cut-off frequencies were set to 1.0–12.0 Hz. Then, the EEG signals were downsampled from 2048 to 32 samples. For *trial extraction,* the following methodology was chosen: from the EEG signals trials of duration 1000 ms were extracted. The starting point of the trial was the stimulus onset, i.e., at the beginning of the intensification of an image, and end point was 1000 ms after the stimulus onset. For *scaling data* a normalization which takes into account the mean and the standard deviation, namely data normalization has been chosen. Several channel configurations have been tested namely:

- 4 electrodes:
- 6 electrodes:
- 8 electrodes:
- 16 electrodes:
- all electrodes (32 electrodes):
- (see Fig. 1)

The pattern construction stage: The samples from the selected channels were concatenated into vectors called EEG classification vectors. The vectors thus created are divided according to the session in two sets, namely, a training set and a test set. For the training set, we also used the information about the illumination of the picture; i.e., we extracted from this set only those EEG vectors that come from illuminating a picture to be watched (that is, a picture that at least theoretically must contain P300). Once these vectors are extracted (i.e., vectors containing P300), we will build a P300 pattern by mediating all of these vectors. The pattern size is given by the number of channels selected, i.e., 32 * Nc, where Nc denotes the number of channels.
Classification stage: The classification stage consists in calculating the scalar product between the P300 pattern built at the previous stage and all the EEG vectors in the test set. We note that several trials were run (between 20 and 25, but in this paper we only used the first 20) and that a trial consists of lighting all 6 pictures. This means that in fact each of the 6 pictures has been illuminated 20 times. Thus, for each picture, in testing stage, we have more values of the scaled products (obtained from scalar product by P300 pattern and EEG signal corresponding to illumination picture), then we add all the values corresponding to each picture and thus obtain a single number, which is actually the sum of all the scalar products of the picture with

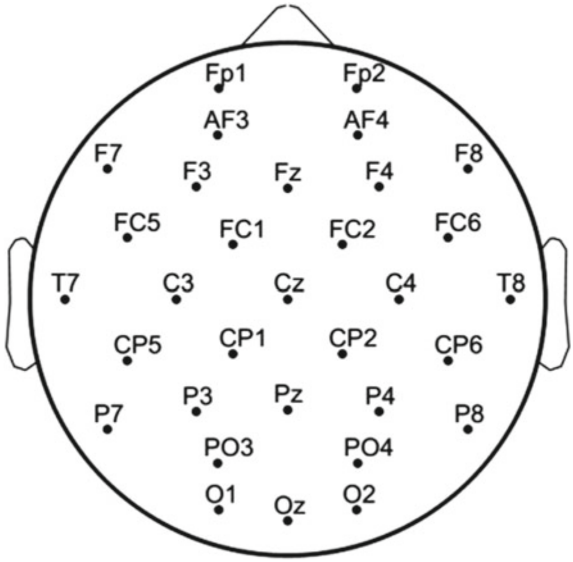


Fig. 1 The configuration with all electrodes

the P300 pattern. Then to classify a picture of the six possible ones, we considered the highest value as belonging to the desired picture.

3 Experimental Results

Figure 2 shows the classification results for all subjects and for all channel configurations tested. It is found that for most subjects, maximum results are obtained when the signals from all channels are used. Subjects 6 and 9 are exceptions, but subject 6 reported that he accidentally focused on the wrong stimulus during one run in session 1, which explains the weaker results. Also, lower performance for subject 9 may be fatigue since somewhat lower performance is restricted to session 4.

There are differences between disabled subjects (1–4) and normal subjects (6–9). Thus, for normal subjects, very good classification rates are obtained even with a smaller number of channels and the classification rate reaches 100%. For subjects with disabilities, the results were much lower than those obtained for normal subjects.

In Table 1, the classification results according to the number of trials are presented. This result was obtained for all channels. In general, increasing the number of trials improves the classification rate. This aspect is also observed if we look at the average values written in the last row of the table.

In Table 2, the average classification for all subjects versus channels configurations for a number of 20 trials is presented. There is a slight increase in the average value with the increase in the number of channels, but it is found that for more than 8

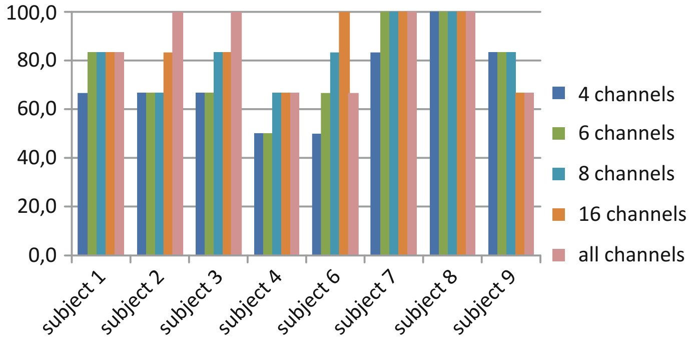


Fig. 2 Classification results versus subjects for several configurations channels for 20 trials

Table 1 Classification results versus subjects for several configurations channels for some trials

No. repetitions					
	5 trials	10 trials	15 trials	20 trials	
Subject 1	66,7	83,3	100,0	83,3	All channels
Subject 2	50,0	100,0	100,0	100,0	All channels
Subject 3	66,7	50,0	83,3	100,0	All channels
Subject 4	33,3	33,3	66,7	66,7	All channels
Subject 6	83,3	66,7	66,7	66,7	All channels
Subject 7	83,3	100,0	100,0	100,0	All channels
Subject 8	100,0	83,3	100,0	100,0	All channels
Subject 9	33,3	66,7	66,7	66,7	All channels
Mean of all subjects	64,6	72,9	85,4	85,4	

Table 2 Average for all subjects versus channels configurations for 20 trials

Average for all subject versus channels configurations	
70,8	4 channels
77,1	6 channels
83,3	8 channels
85,4	16 channels
85,4	all channels

channels the growth is small and it is not justified to increase the number of channels to increase calculation time or the necessary resources.

4 Conclusions

This paper has presented a new and efficient P300 detection method of BCI system. With the proposed method good results for a large number of test trials are obtained. It has been found that the increased number of channels helps to improve the results. The accuracy classification obtained is 85% using EEG signals on all channels and 20 testing trials.

Acknowledgements We want to thank all human subjects who have voluntarily participated in experiment and Ulrich Hoffmann and his team for permission to use the EEG data available on the Internet.

References

1. Birbaumer N, Ghanayim N, Hinterberger T, Iversen I, Kotchoubey B, Kübler A, Perelmouter J, Taub E, Flor H (1999) A spelling device for the paralyzed. Nature 398:297–298
2. Mason SG, Birch GE (2000) A brain-controlled switch for asynchronous control applications. IEEE Trans Biomed Eng 47:1297–1307
3. Pfurtscheller G, Flotzinger D, Kalcher J (1993) Brain-computer interface: a new communication device for handicapped persons. J Microcomput Appl 16:293–299
4. Wolpaw JR, McFarland DJ, Neat GW, Forneris CA (1991) An EEG-based brain-computer interface for cursor control. Electroencephalogr Clin Neurophysiol 78:252–259
5. Jones KS, Middendorf M, McMillan GR, Calhoun G, Warm J (2003) Comparing mouse and steady-state visual evoked response-based control. Interact Comput 15:603–621
6. Farwell LA, Donchin E (1988) Talking off the top of your head: a mental prosthesis utilizing event-related potentials. Electroencephalogr Clin Neurophysiol 70:510–523
7. Donchin E, Spencer K, Wijesinghe R (2000) The mental prosthesis: assessing the speed of a P300-based brain-computer interface. IEEE Trans Rehab Eng 8(2)
8. Sutton S, Braren M, Zubin J, John ER (1965) Evoked correlates of stimulus uncertainty. Science 150:1187–1188
9. Donchin E (1980) Presidential address. Surprise!…Surprise? Psychophysiology 18:493–513
10. Fabiani M, Gratton G, Karis D, Donchin E (1987) Definition, identification and reliability of the P300 component of the event-related brain potential. In: Ackles PK, Jennings JR, Coles MGH (eds) Advances in psychophysiology, vol 2. JAI Press, New York, pp 1–78
11. Polich J (1999) P300 in clinical applications. In: Niedermeyer E, Lopes da Silva FH (eds) Electroencephalography: basic principles, clinical applications and related fields, 4th edn. Williams and Wilkins, Baltimore, pp 1073–1091
12. Rosenfeld JP (1990) Applied psychophysiology and biofeedback of event-related potentials (brain waves): historical perspective, review, future directions. Biofeedback Self Regul 15:99–119
13. Coles MGH, Rugg MD (1995) Event-related potentials: an introduction. In: Rugg MD, Coles MGH (eds) Electrophysiology of the mind: event-related brain potentials and cognition. Oxford University Press, New York
14. Kramer AF, Strayer DL (1988) Assessing the development of automatic processing: an application of dual-track and event-related brain potential methodologies. Biol Psychol 26:231–267
15. O'Donnell BF, Friedman S, Swearer JM, Drachman DA (1992) Active and passive P3 latency and psychometric performance: influence of age and individual differences. Int J Psychophysiol 12:185–187

16. Polich J (1986) Attention, probability, and task demands as determinants of P300 latency from auditory stimuli. Electroencephalogr Clin Neurophysiol 63:251–259
17. Hoffmann U, Vesin JM, Ebrahimi T, Diserens K (2008) An efficient P300-based brain–computer interface for disabled subjects. J Neurosci Methods 167(1):115–125 (15 Jan 2008)
18. http://infoscience.epfl.ch/record/101093

Detection of Weed Using Visual Attention Model and SVM Classifier

Manda Aparna and D. Radha

Abstract Agriculture is one of the provenances of human ailment in this heavenly body. It plays an extrusive role in the economy. Flourishing crops are a constituent of agriculture. Weeds are the additional plants to the crop. Removal of weeds is a challenging job for the farmers as it is a periodic, time–consuming, and cost-intensive process. Different ways to remove those weeds are by hand labor, spraying pesticides and herbicides, and machines but with their own disadvantages. The software solution can overcome these drawbacks to an extent. The main concern in software is in the identification of weeds among the crops in the field. The proposed system helps in detection of weeds in the agriculture field using computer vision methods. The method works with a dataset of crops and weeds. The plants are identified as salient regions in visual attention model and the identified plants are classified as crops or weeds using support vector machine classifier.

1 Introduction

Weed infestation is a universal complication in agriculture that negatively influences crop yielding. Despite, the method of herbicides constantly in a farmland has concluded in serious environmental contamination. The utmost trivial measures are definiteness spraying with choosy pesticides to preserve the assurance of drinking water and to diminish environmental collision in agriculture. Researchers have become increasingly vigilant of the crucial lead precision agriculture (PA) will play in the future [1]. Due to weeds random property, discrete approaches and methods for automatic weed recognition by machine perception have been proposed. Machine perception for discriminatory weeding or choosy herbicide sprinkling completely

M. Aparna · D. Radha (✉)
Department of Computer Science & Engineering, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Bengaluru, India
e-mail: d_radha@blr.amrita.edu

M. Aparna
e-mail: mandaaparna76@gmail.com

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_25

confides on the capability of the ideology to evaluate weed images. Despite the image, segmentation may be favored as the presently primary step for detecting the objects using machine vision. Its unsteadiness arising from instability of the outdoor operating environment is a central interference to influence precisely the instability and rapidly abstract region-of-interest features from the immense measure of in-field image evidence is a stimulating complication.

The human's visual entity has a visual attention appliance that assists individuals promptly which prefers the vastly significant enlightenment from a scene. Featuring visual attention and simulating eye saccades are very profitable measures in computer vision. Recently, advancing reckoning visual attention representations are used to replicate the human visual appliance has been inducing enhanced interest in the domain of computer vision. Based on this, the foremost goal of this analysis is to detect plant as salient regions using visual attention mechanisms and computer vision theory in the outdoor operating environment and differentiate them as weed and crop.

For differentiating them as weed and crop here, we are using support vector machine as a classifier. Support vector machine (SVM) is a superintended machine learning innovation, which can be pre-owned for both classification and regression confrontations. Mainly it is issued for the classification process.

2 Related Work

In earlier days, the most common technique of removing weed in the crop was by manual monitoring in which a man was separately appointed for cross-checking of weed and crop and removing it by hand in the case if it was a weed. Nowadays, disparate image processing approaches have been adopted for the removal of weed in the agricultural farmland. Some previous works are listed as below. In earlier days, weeds were identified by simple image processing methods. The techniques involved were texture-based weed detection algorithm, probabilistic neural networks [PNN] and many more techniques were involved which was complex and time-consuming. In this algorithm, greenness identification and morphological operations were involved. This morphological operation considers many mathematical operations. This classification algorithm contains noise [2].

Weed detection is done using image processing and clustering analysis. It involves edge, texture properties, and clustering analysis. The results are not convincing in which the edges are not well defined. It is computationally expensive. Non-globular clusters [3] are not suitable in this case.

Other methods used for weed detection are Morphological functions, Otsu thresholding and Artificial Neural Networks (ANN). Here, morphological operations are used only for segmentation. In Otsu thresholding method, the performance degrades, when the global distribution of target image and background varies widely. The network structure of ANN is complex comparatively [4].

Image processing methods like canny edge detection, threshold algorithm, and morphological operations are also used for weed detection. This was a time-consuming process and selection of threshold is crucial because the wrong selection may result in over or under-segmentation [5]. In automated weed assortment including local pattern-relied texture descriptors, they used texture-based classification method and the classifier support vector machine (SVM). In this, the result is not convincing, as there is no preprocessing and because of it, noise is generated. A massive number of support vectors is desired from the training firm to execute regulation responsibility [6].

In weed detection system, the consequence of automatic systematization of broad and narrow weed utilizing characteristic vector were extricated by a composition of Gabor filter and FFT, and the classification employing the support vector machine (SVM). The weed classification routine comprises three processes like pre-transforming, feature extrication, and executing. Different parameters of SVM have to be analyzed for getting better results and a large number of images in the dataset have to be used for training [7]. The other way is vision-planted classifier in which Bayes and support vector machines, image segmentation, and decision-making are involved in detection of weed. The similarity in the texture and spectral signatures of the cereal crop and the weed leads to difficulty in weed detection [8].

The proposed system uses the visual attention model, which makes the salient region to be highlighted, and the classification is based on the salient region. The background and foreground are separated and classification is performed using support vector machine. This system works well in classifying broadleaf plants as crops and narrow leaf plants as weeds.

3 Proposed System

3.1 Visual Attention Model (VAM)

In modern lifetime, analyzers have determined that adopting human visual attributes to disclose objective could gain affirmative conclusions. The judicious attention property of human vision benefits inhabitants to pursue the salient object instantly and accurately from desolated locality without getting affected by complicated environment. Ultimately, the current saliency model structure usually needs an aggregation of the elementary visual features to propagate saliency maps, such as intensity, orientation and color information, and so on. This map provides a precise aspect of the visual environment, affirming salient locations in the visual field. The preference of visual features for saliency detection relies upon the evident application. Whereas assorted visual features contribute a discrete improvement to saliency detection, an assured feature may be capable in one case but fragile in another. Providing an input image, the first processing step consists of dissolving this input into an assortment of the apparent features: namely intensity, color, and orientation. The feature maps are created for various features as shown in the Fig. 1.

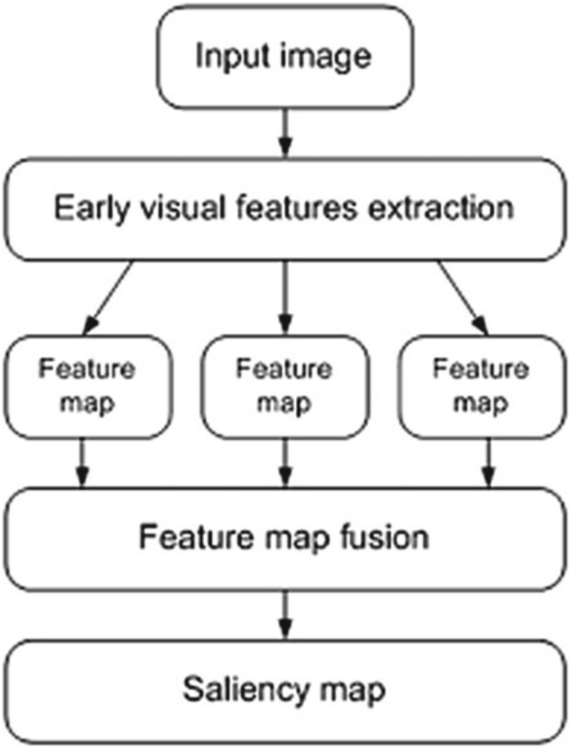


Fig. 1 Saliency mapping in visual attention model

3.2 Support Vector Machine (SVM)

The SVM approach is a current literature setup for coordination employment. Support vector machine (SVM) is an approach for constructing an optimal binary (2-class) classifier. Figure 2a displays a dual-class issue by multifold available hyperplanes splitting the pair data sets that are not naturally superlative. In Fig. 2b, an ideal splitting hyperplane (OSH) is exhibited which produces the utmost limit (dashed border) intervening the pair of data sets. SVM discovers this OSH by inflating the limit among the classes. SVM initially converts prescribed information into a superior dimensional zone by path of a kernel function and then builds a linear OSH among the pair of classes in the commuted slot. Those evident vectors adjacent to the established line in the transfigured field are labeled as the support vectors (SV). SVM is a relative employment of the approach of "structural risk minimization" addressing to gain subsided possibility of principle inaccuracy. To elevate SVMs, a kernel objective is preferred [9].

In this analysis, the radial basis function (RBF) kernel [10] is preferred for selection of SVM parameter. RBF kernel is ultimately utilized to overcome SVM

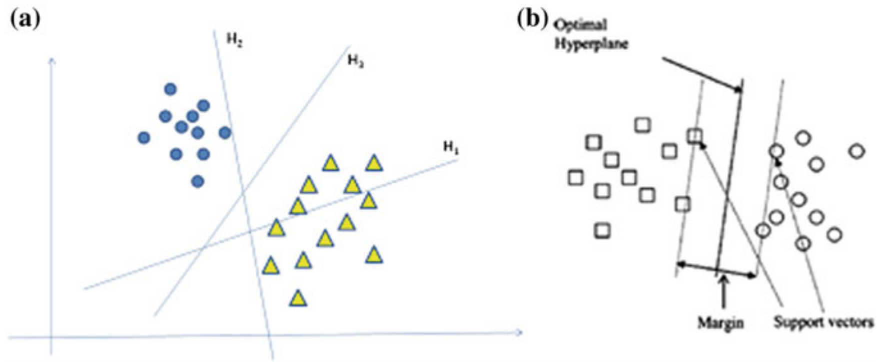


Fig. 2 **a** Abundant feasible distributing hyperplanes isolating the couple of groups. **b** Ideal dividing hyperplanes and the ultimate border. The circles and squares exhibit elements of two groups

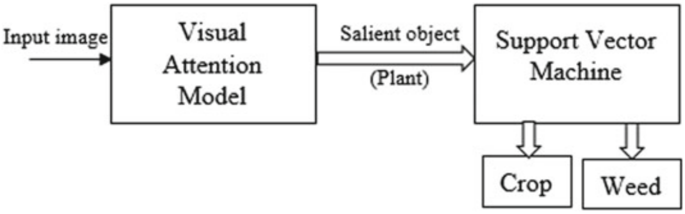


Fig. 3 Suggested system

classification issues. The generalized representation of the suggested system consists of the computational part as displayed in Fig. 3.

In the prospective system, the input image is given to the visual attention model, which results in extracting the saliency map of the image. This saliency map undergoes binarization that helps in locating the shape of the region of interest as a plant. Finally, the support vector machine classifies the plant into weeds or crops.

4 Implementation

The implementation consists of three modules namely visual attention model, binarization, and support vector machine. The plant image is given as input to the visual attention method. In visual attention, several processes are involved to abstract the region of interest needed. The primary process involved is feature extrication [11]. Feature extrication is a type of scale degradation that smoothly imitates the captivating element of an image as a solid attribute point. This access is appropriate when an image immensity is enormous and an abbreviated feature embodiment is recommended to instantly finish tasks such as image matching and image retrieval. The next process is saliency mapping. In computer vision, a saliency map is an image that displays each pixel's individual characteristic. The intention of a saliency map [12] is to elucidate and/or compressing the delegation of an image into the commodity that is more purposeful and effortlessly to estimate. The objective of saliency map is to embody the eminence or saliency in whole region of visual field by a scalar measure. It is used to lead the choosing of the attended region depending on the geographical allocation. The aggregation of feature maps supports bottom-up input to the saliency map.

Binarization is similar to segmentation for the identification of salient object from salient region. The main goal of binarization is separation of background and foreground of the plant image [13]. It is processed by using threshold. By using this threshold, the salient region of the plant image is obtained [14].

Finally, the classification of the plant object as crop or weed is identified by SVM classifier from the dataset. SVM classifier is trained with the training dataset to acquire needed splitting hyperplanes.

The dataset considered for the analysis are manually collected images with 31 broadleaf images and 47 narrow leaf images [15]. Broadleaf images are taken as crops and narrow leaf images are considered as weed. 55 images are randomly selected for training module and 31 broadleaf images, and 47 narrow leaf are used for the testing module. The results obtained from the training dataset are used to test the images and classify it as crop or weed. The implementation is shown in Fig. 4.

5 Results and Analysis

The plant image is given as input to visual attention model in order to obtain saliency map, binaralized saliency map, and salient region. Saliency map of input plant image exhibits each pixel's exclusive characteristic. Saliency mapping is a set of contours extricated from the image. Each of the pixels in the region is identical relative to some qualities such as color, intensity, and orientation. The objective of saliency map is to elucidate an input image into an entity for easier analysis. Binaralized saliency mapping is computed in which the foreground and background are shown in different colors. It is used to detect boundaries of an image. Saliency maps are composed at various scales. These maps are combined pixel-wise to achieve the eventual saliency map. The salient region of plant image is resolved as the local contrast of an image region relative to its surrounding at different scales. Finally, the salient region is detected. Figure 5 shows the computation process of the visual attention model.

The SVM classifier is used for identification of weeds and crops. The training and testing images are in a predefined folder. Here, 55 images are used for training and 78 images are used for testing. The Gabor kernel and Fourier transform (FT) carry out the filtering of these images and input real-time image. Filtering is a method for altering or embellishing an image. Gabor kernel filter is a linear filter applied for texture analysis which fundamentally determines any precise frequency content in the image in any particular directions in a bounded region around the analysis region. The Fourier transform (FT) is a crucial image processing tool, which is used

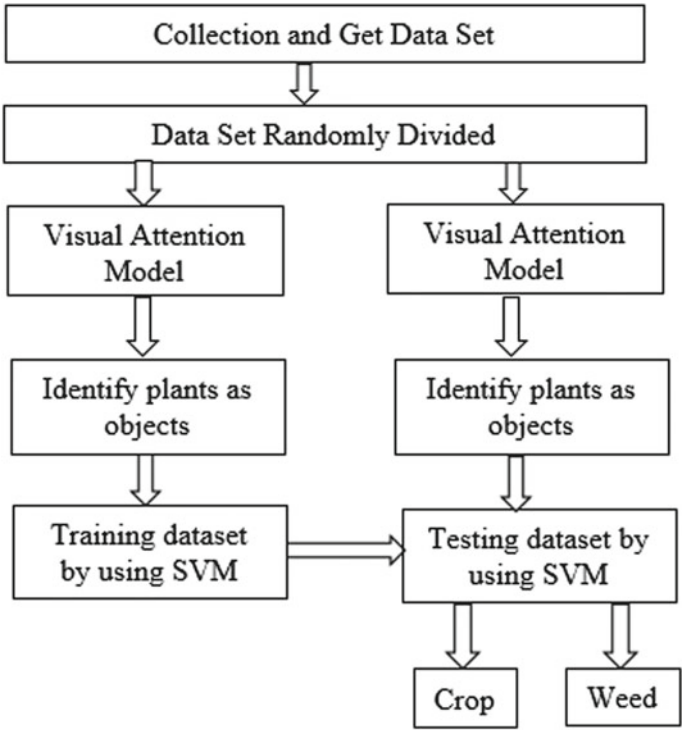


Fig. 4 SVM testing and training modeling framework

Table 1 Results of crop input images

Total number of crop images	Number of crop images detected as crops correctly	Number of crop images detected as weeds (Incorrect)	Percentage of detecting crops correctly
31	20	11	64.5

to dissolve an image into its sine and cosine parts. In the Fourier, concern image, each point shows specific frequency contained in the spatial domain image. In this paper, assumption of weed is by 0 and crop is by 1. All these values of images are stored in an array. SVM classifier performs the identification of weed and crop from input image by comparing the features of the training images. Figure 6 is a sample crop image given as input for visual attention model and the SVM classifier for crop or weed identification. Figure 7 is a screenshot of the result obtained for the input image shown in Fig. 6.

A sample weed image shown in Fig. 8 as input for visual attention model and the SVM classifier for crop or weed identification. Figure 9 is a screenshot of the result obtained for the weed input.

The results of crops images as input is shown in Table 1.

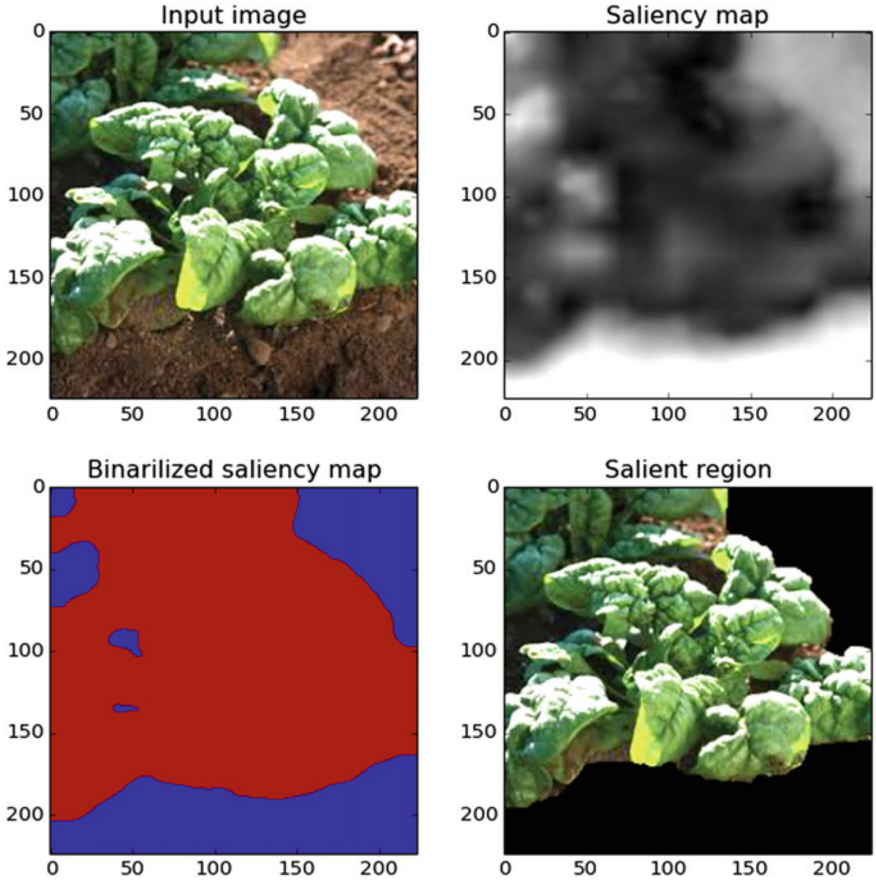


Fig. 5 Computation process

Fig. 6 Crop image

The results of weeds images as input is shown in Table 2.

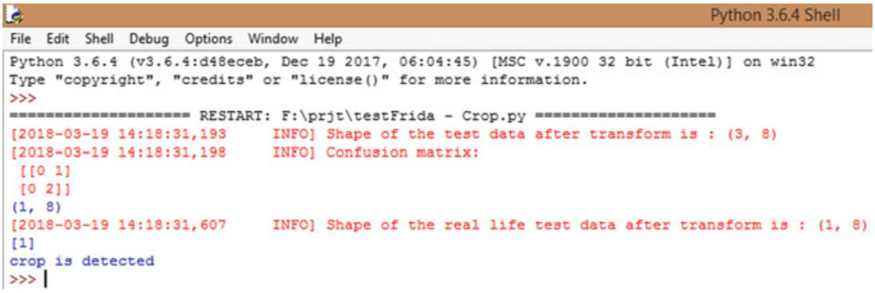


Fig. 7 Detection of crop

Fig. 8 Weed image

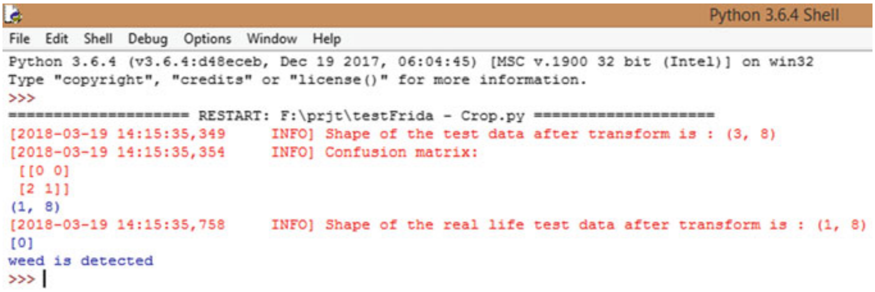


Fig. 9 Detection of weed

6 Conclusion

Removal of weeds is a challenging job for the farmers as it is a periodic, time—consuming, and cost-intensive process. The competence of spotting and analyzing crops and weeds could spark the evolution of sovereign vision-instructed agricultural

Table 2 Results of weed input images

Total number of weed images	Number of weed images detected as weeds correctly	Number of weed images detected as crops (Incorrect)	Percentage of detecting weeds correctly
47	40	7	85.1

contraptions for site-peculiar herbicide employment. In this paper, we have suggested a categorization type model which uses visual attention model and support vector machine(SVM) to organize the plants in the images as crop or weed in order to scale down the exaggerate treatment of herbicides in agricultural systems. Analysis of the result exposes that the SVM acquires good percentage of accuracy in identifying the weeds. This type of classification can be used for the agricultural robot to distinguish the crops and weeds which in turn can help in removal of weeds.

References

1. Wu L (2017) Detection of the salient region of in-field rapeseed plant images based-on visual attention model. In: 2017 2nd Asia-Pacific conference on intelligent robot systems (ACIRS). IEEE, New York, pp 33–36
2. Mohanapreethi J, Sujaritha M (2017, May) Texture-based weed identification system for precision farming. Int J Sci Eng Res 8(5). ISSN 229-5518
3. Sukumar P, Ravi S (2016) Weed detection using image processing by clustering analysis. Int J Emerg Technol Eng Res (IJETER) 4(5):14–18
4. Ayswarya R, Balaji B, Balaji R, Ramya R, Arun S (2017, March) Weed detection in agriculture using image processing. Int J Adv Res Electr, Electron Instrum Eng 6(3)
5. Deepa S, Hemalatha R (2015) Weed detection using image processing. Int J Res Comput Appl Robot 3(7):29–31
6. Ahmed F, Kabir MH, Bhuyan S, Bari H, Hossain E (2014) Automated weed classification with local pattern-based texture descriptors. Int Arab J Inf Technol 11(1):87–94
7. Ishak AJ, Mustafa MM, Tahir NM, Hussain A (2008) Weed detection system using support vector machine. In: International symposium on information theory and its applications, 2008, ISITA 2008. IEEE, New York, pp 1–4
8. Tellaeche A, BurgosArtizzu XP, Pajares G, Ribeiro A (2007) A vision-based classifier in precision agriculture combining Bayes and support vector machines. In: IEEE international symposium on intelligent signal processing, 2007, WISP 2007. IEEE, New York, pp 1–6
9. Auria L, Moro RA (2008) Support vector machines (SVM) as a technique for solvency analysis
10. Ahmed F, Bari ASMH, Hossain E, Al-Mamun HA, Kwan P (2011) Performance analysis of support vector machine and Bayesian classifier for crop and weed classification from digital images. World Appl Sci J 12(4):432–440
11. Venkataraman D, Mangayarkarasi N (2016) Computer vision based feature extraction of leaves for identification of medicinal values of plants. In: 2016 IEEE international conference on computational intelligence and computing research
12. Sriram B, Sai Hemanth Reddy K, Santhosh Kumar S, Sikha OK (2017) An improved levelset method using saliency map as initial seed. In: International conference on signal processing and communication (ICSPC'17)—28th & 29th July 2017
13. Malemath VS, Hugar SM (2016, May-June) A new approach for weed detection in agriculture using image processing techniques. Int J Adv Sci Tech Res 3(6). ISSN 2249-9954

14. Cheng M-M, Warrell J, Lin W-Y, Zheng S, Vineet V, Crook N (2013) Efficient salient region detection with soft image abstraction. In: 2013 IEEE international conference on computer vision (ICCV). IEEE, New York, pp 1529–1536
15. Anjali Rani KA, Supriya P, Sarath TV (2017) Computer vision based segregation of carrot and curry leaf plants with weed identification in carrot field. In: Proceedings of the IEEE 2017 international conference on computing methodologies and communication (ICCMC)

Design and Development of Scalable IoT Framework for Healthcare Application

Siddhant Mukherjee, Kalyani Bhole and Dayaram Sonawane

Abstract With increasingly fast-paced life and alarmingly high rate of chronic ailments in general population, there is a need for quickening the current process of healthcare monitoring, especially in emergency situations. The recent developments in communication technology, especially in the field of Internet of Things (IoT) have enhanced the accessibility of such systems. This can be achieved by transmitting/uploading the data of various health parameters acquired by different physiological sensors with wireless sensor networks onto the cloud platform. This data later can be accessed by the concerned medical authorities when required for diagnosis. In this work, we focus on the development of the scalable IoT framework for monitoring the physiological parameters of the patient. The customized MATLAB-based GUI is designed to perform real-time analysis of sensor's data which is used for continuous monitoring of vital parameters of the body. We have developed a wearable band which can be worn as a wristband by the patient. The band consists of temperature, pulse, and ECG sensors those are used to transmit the vital parameters of the patient integrated with the ultra-low-power battery-operated Texas Instruments MSP430 microcontroller with CC110L sub-1 GHz RF wireless transceiver. The concept is successfully demonstrated by transmitting three physiological parameters wirelessly over 100 m distance as well as over the cloud platform.

1 Introduction

Wireless sensor networks are growing as a platform which enhances the relationship between the devices used by several communities for generations and devices which constantly make the use of Internet and maintain the organization of data related to

S. Mukherjee · K. Bhole · D. Sonawane (✉)
College of Engineering Pune, Shivajinagar, Pune, India
e-mail: dns.instru@coep.ac.in

S. Mukherjee
e-mail: mukherjeesg16.instru@coep.ac.in

K. Bhole
e-mail: kab.instru@coep.ac.in

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_26

showcasing the condition of the environment through means of the physical parameters collected by a single station. Applications vary from the use of such networks to highlight problems related to a particular societal cause ranging from manufacturing and health care to retail business and home automation. Wireless sensor networks or WSNs are budding technology with an immense amount of potential to change the way people are living their lives. WSNs have been used to enable better data collection in scientific studies, create more effective strategic military defenses and monitor factory machinery [1]. Wireless sensors can be classified on the basis of readiness for field deployment, scalability, and cost. Readiness for field deployment measures maturity for field deployment in terms of economic and engineering efficiency. The requirement for the sensor to be scalable to distributed environmental monitoring tasks requires that the sensors are small and inexpensive and architecture needs to be equally capable enough to accommodate or scale-up to many distributed systems. Sensors are deployed in large numbers and a prerequisite for the deployment of such sensors is that cost will drop and also energy requirement for maintaining the healthy state of a scalable architecture incorporating such sensors should be low. Sensor networks represent a significant advantage over traditional invasive methods of monitoring and also represent a more economical method for conducting long-term studies than traditional methods [2].

Sensor networks used for medical applications mostly range their use either as an implantable or as a wearable. Implantables can be inserted into a human body, whereas wearables are devices that are used to be worn on the surface of the skin or to the nearest portion of the body. Body area networks can collect information about an individual's health, fitness, and energy expenditure [3, 4]. Scalability is a very important and crucial issue in the design of routing protocols for WSNs. Scalability involves sensitivity and range of sensors, communication bandwidth of the radio and power usage. The software issues include reliability of data transfer, development of a methodology to maintain the management of a large volume of data and scalable algorithms for viewing and analyzing the data in real time. However, there needs to be a proper balance between the increased number of nodes and the rise in complexity to manage such increase in nodes. Internet of things or IoT is a new Internet paradigm based on the fact that there will be many more things than human connected to the Internet.

Jiao et al. (2017) adopted CC2420 transceiver chip which used 2.4 GHz radio frequency band for wireless transmission of sensor data to the remote site [5]. We have used CC110l transceiver chip which supports 915–920 MHz radio frequency bands in this project for communicating sensor values between two transceivers. However, CC110l consumes less receiver current (i.e., 14 mA) compared to CC2420 (i.e., 12 mA). Gupta et al. (2010) designed a MATLAB GUI for acquisition and online analysis of ECG signal along with an embedded system [6]. In this project, Matlab GUI was used for acquisition and analysis of temperature, pulse, and ECG sensor at the same time as it helps in real-time analysis of raw data coming from these sensors, i.e., one can understand if the patient is suffering from cardiac arrhythmia by analyzing his/her QRS complex with respect to time interval in milliseconds at which the signal is recorded. Moreover, one can switch quickly among various buttons to visualize sensor data separately or all at the same time which indeed offers an instantaneous view of physical parameters of the body.

This paper introduces the development of such framework for three sensors which can be scaled up to eight sensors depending upon the application. The system will make use of local RF band for transmission and reception of sensor data and the data will be sent to a web-based dashboard using Wi-Fi protocol. This paper also mentions the use of MQTT as an application layer to send the data and its advantages over other protocols. This paper also presents a real-time user interface developed in Matlab to easily monitor sensors irrespective of their number and represents the data graphically as well as numerically and simultaneously store data onto a file which keeps a record of such data. The paper includes the following sections: Sect. 2 provides detailed information about system architecture, Sect. 3 highlights results obtained after implementation of the system architecture and incorporating sensors with it and discussions upon it, and Sect. 4 offers conclusions and future scope of this project along with references.

2 System Architecture

The system consists of blocks that represent the whole architecture which is as follows:

- Sensor Block
- Transmitter Block
- Receiver (Gateway) Block
- Data Acquisition Block.

The Transmitter Block communicates with Receiver (Gateway) Block via local RF protocol and Receiver (Gateway) Block transfers such data to the web-based dashboard over Wi-Fi which can be viewed on a personal computer or on any device that supports Wi-Fi connectivity. This architecture enables any sensor connected to the transmitter block to communicate with the central station which supports Wi-Fi connectivity independently. Figure 1 represents the system architecture, the figure also shows a connection flow among different components and how they communicate with each other.

2.1 Sensor Block

The Sensor Block consists of sensors which are used for monitoring the health status of an individual. Sensors used for this purpose are temperature sensor, pulse sensor, and ECG electrodes. LM35 is used to sense and measure the temperature of the body. LM35 is a temperature sensor IC which measures temperature in the range of 0–100 °C which also covers body temperature range of 4–40 °C and is calibrated

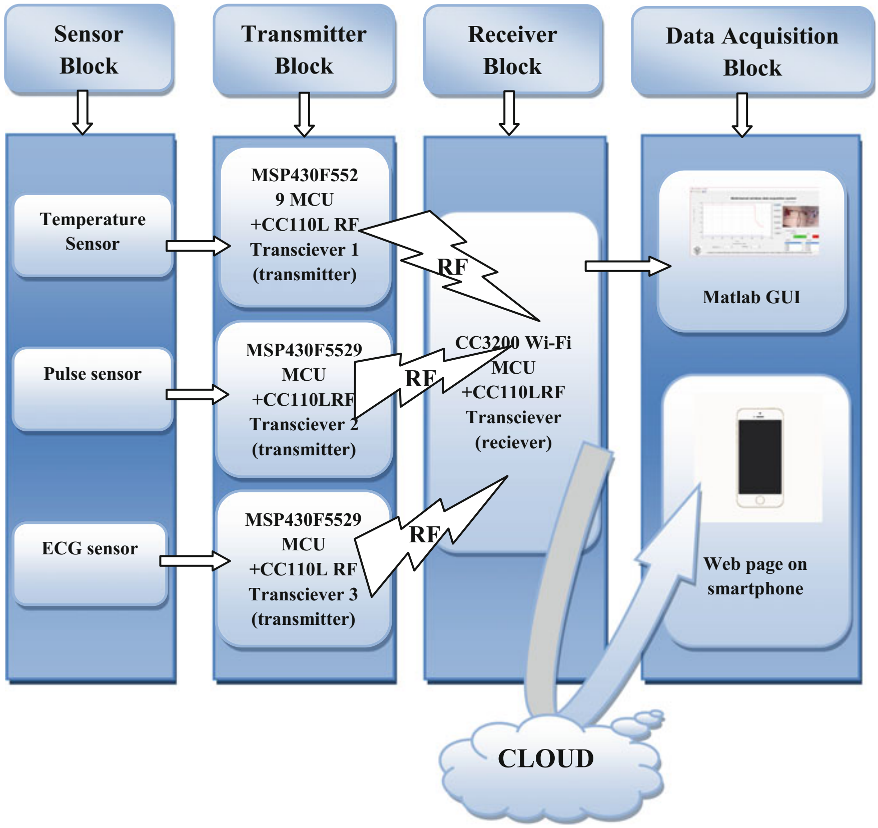


Fig. 1 Block diagram of a proposed system architecture

and connected to an analog channel of MSP430F5529 microcontroller unit as shown in Fig. 2. Pulse sensor is used to calculate the BPM or Beats Per Minute which is connected to the analog channel of separate MSP430F5529 microcontroller unit as shown in Fig. 2. Pulse sensor unit consists of a green led and a low-cost ambient light photosensor (photodiode) and placed at the tip of the finger or near the veins on an arm. Pulses are calculated via reflective method, i.e., light from green led falls on the surface and gets reflected as soon as the pulse is detected (blood cells are pumped into the blood vessels and forms a pulse). BPM is calculated by calculating the difference between the peak of two consecutive pulses and multiplying by 60 s. ECG electrodes were used as the third sensor to measure ECG by placing three Ag–AgCl electrodes at the right arm, left arm and right leg, respectively. These electrodes were connected to the AD8232 analog front-end ECG module which consists of analog high-pass filter and low-pass filter to remove the noise from ECG signal which are presented as a result of 50 Hz baseline interferences. The AD8232 module was connected to the analog channel of the third MSP430F5529 microcontroller unit as shown in Fig. 2.

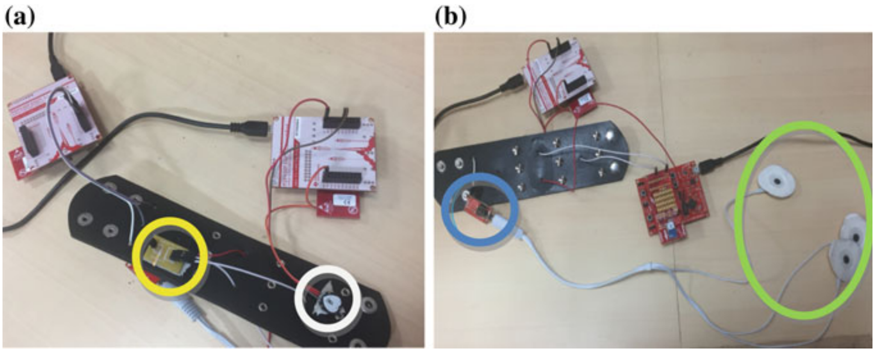


Fig. 2 Health band showing **a** temperature sensor encircled in yellow and pulse sensor encircled in white and **b** AD8232 ECG module encircled in blue and ECG electrodes encircled in yellow

2.1.1 MSP430 and Its Applications in Health Care

Devices used for monitoring personal health requires the availability of some of the key features like ultra-low-power consumption, battery life which decides form factor, reduced size, and so microcontrollers need to be chosen accordingly. The main reason behind using MSP430 for healthcare applications is that it is designed for an ultra-low-power application because it utilizes a flexible clocking system and a variety of operating modes designed to reduce power consumption and hence, extending battery life which can be said by the fact that the current drawn during active mode is in the range of 100–200 μA/MHz and less than 1 μA during a standby mode and can also operate down to 1.8 V, further improving a superior power specification [7]. One can achieve a 10-year life with 1 Ah battery using MSP430 as it provides 10 μA average power budget for the entire system which can be useful for developing both portable as well as implantable personal healthcare devices [8].

2.2 Transmitter Block

The Transmitter Block consists of Texas Instruments MSP430F5529 MCU or microcontroller unit to which Anaren's CC110L Air Transceiver Booster Pack is attached. CC110L Air Transceiver Booster Pack is connected to MSP430F5529 MCU via SPI interface and is programmed as a transmitter. This transmitter sends the relevant sensor data in the form of RF packets and uses sub-1 GHz RF, i.e., 915–920 MHz. Sub-1 GHz has several benefits over 2.4 GHz radios of being low power and give coverage up to 1–4 km and can pass through walls much better than 2.4 GHz [9]. In this project, three such transmitters to which three sensors were connected are shown in Fig. 2a, b. However, a number of sensors connected to the system can be increased or scaled up to eight sensors depending upon the application.

2.3 Receiver (Gateway) Block

The Receiver (Gateway) Block consists of Texas Instruments CC3200 Wi-Fi Microcontroller Unit which features Wi-Fi CERTIFIED™ chip and uses IEEE 802.11 bgn which means it supports 900 MHz–2.4 GHz single band frequency band [10]. CC3200 Wi-Fi MCU also comes with Enhanced IoT Networking Security and has been optimized for low power. In this project, CC110L Air Transceiver Booster Pack is connected to CC3200 Wi-Fi MCU via SPI interface and programmed as a receiver. This receiver accepts the RF packet sent by the transmitter depending upon the pre-requisite that transmitter and receiver should have the same channel ID and the same initial address. The received sensor packet from the receiver goes to the separate 12-bit ADC channels of CC3200 Wi-Fi MCU chip and such data can be acquired by gaining access to the serial port of CC3200 Wi-Fi MCU. As sensor data can be accessed from the ADC channels, we can increase the number of sensors connected to Transmitter Block depending upon ADC channels present in CC3200 Wi-Fi MCU which makes this IoT architecture scalable and more efficient to be used.

2.4 Data Acquisition Block

The Data Acquisition Block comprises a real-time based user interface which is utilized to visualize the collected data by gaining access to the serial port of CC3200 Wi-Fi MCU. MATLAB GUI was designed which can connect to the serial port of the CC3200 Wi-Fi MCU and allows visualization of data graphically as well as numerically and simultaneously store data on the folder onto which MATLAB GUI was made and also consist of different channels which can show the individual status of sensors and their data. Further, CC3200 Wi-Fi MCU was programmed to connect to a Wi-Fi network by adding credentials of that Wi-Fi network and send data to the web-based dashboard using MQTT protocol as an application layer. MQTT or Message Queue Telemetry Transport is a publish/subscribe protocol that runs on the top of the transport layer. This protocol is better for IoT as it provides lower network bandwidth and increases the lifetime of battery-run devices [11]. MQTT protocol requires the name of the topic under which sensor data will be sent to the MQTT broker and data can be accessed on a web-based dashboard which can be seen from anywhere on an internet webpage via PC or smartphone having excellent Internet connectivity.

3 Results and Discussions

Results are obtained after implementing the architecture as shown in Fig. 1 and sensor data from health band were visualized on MATLAB GUI as well as on web-based dashboard shown in Figs. 3, 4, 5 and 6.

Results obtained from sensors via data acquisition using MATLAB GUI shows the freedom of visualizing important medical information from data by clicking on any one of the three channels and analyzing all such information in real time. We have also learned that instead of relying on patient monitors and physical presence of a physician for the analysis of physical parameters of a patient, a physician can continuously monitor the patient's health from his smartphone having internet connectivity and give his/her expert advice to the patient from anywhere.

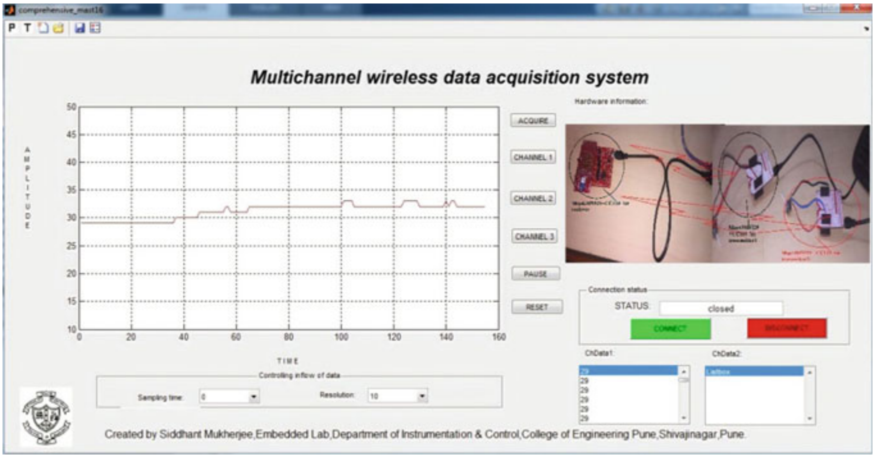


Fig. 3 Temperature sensor showing the temperature of the body in MATLAB GUI

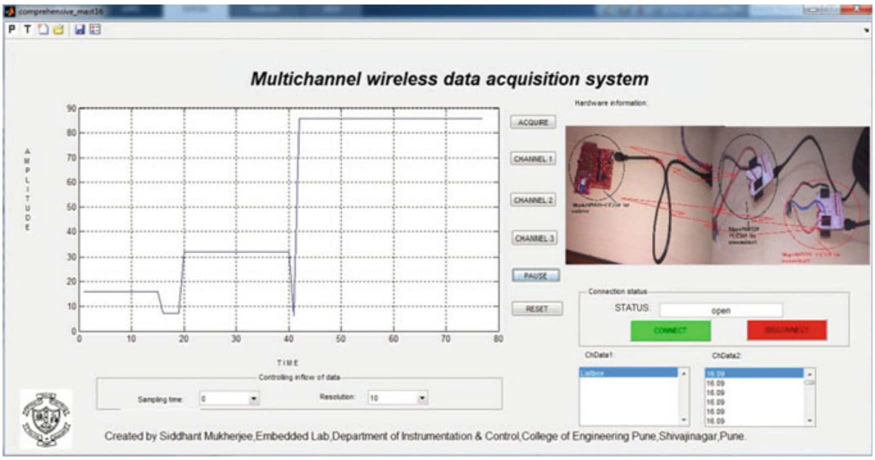


Fig. 4 Pulse sensor showing BPM of the person in MATLAB GUI

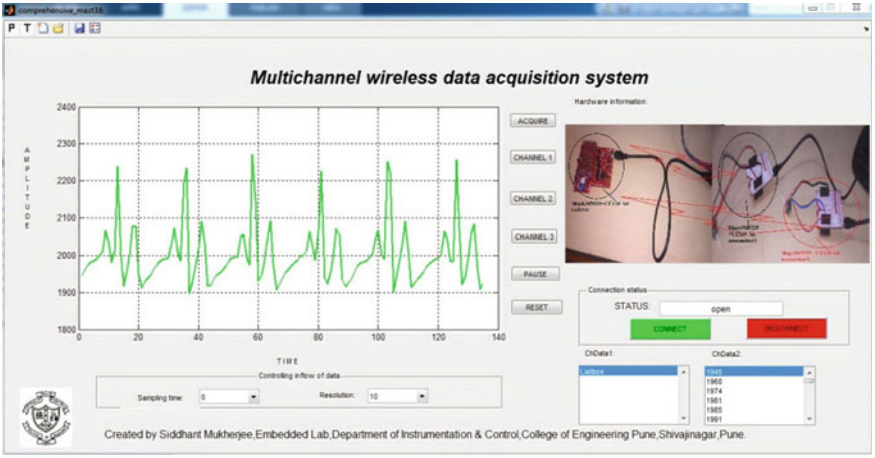


Fig. 5 ECG electrodes showing ECG of the person in MATLAB GUI

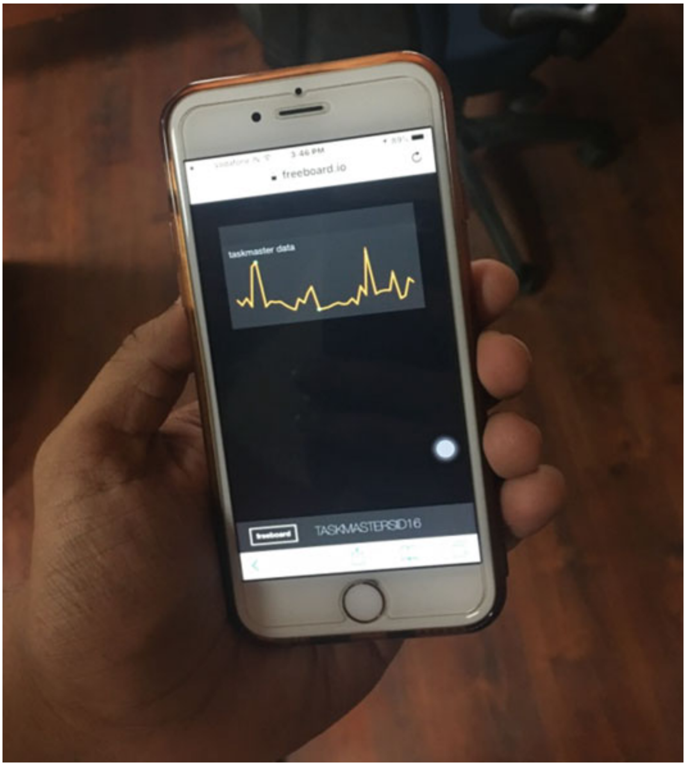


Fig. 6 Sensor data displayed on a webpage on an IOS-based smartphone

4 Conclusions and Future Scope

Data from body temperature, pulse, and ECG sensors were collected and visualized using MATLAB GUI and webpage and the IoT framework for wireless sensor network designed for healthcare application (health band) was successfully tested for its scalability, i.e., ability to accommodate utilization of around eight sensors. In the future, we have planned to work on power optimization to extend battery life and increase the durability of the wristband. It is also planned to incorporate two-finger ECG onto a wristband as well as extraction of the features derived from the received physiological parameters of a patient. We look forward to visualizing the possibility of coming up with the commercialized model of a wristband having three sensors which will be communicating the vital parameters of a patient over local wireless networks as well as over the cloud platform.

References

1. Culler DE, Mulder H (2004) Smart sensors to network the world. Sci Am 290:52–59
2. Mainwaring A, Polastre J, Szewczyk R, Culler D, Anderson J (2002) Wireless sensor networks for habitat monitoring. In: WSNA ’02 Proceedings of the 1st ACM international workshop on wireless sensor networks and applications, pp 88–97
3. Peiris V (2013) Highly integrated wireless sensing for body area network applications. SPIE Newsroom
4. O’Donovan T, O’Donoghue J, Sreenan C, Sammon D, O’Reilly P, O’Connor KA (2009) A context-aware wireless Body Area Network (BAN) (PDF). In: Pervasive computing technologies for healthcare
5. Jiao C, Cheng G, Tang M, Chen S (2016) Portable monitoring instrument of the physiological parameter based on MSP430 microcontroller. In: 2016 IEEE advanced information management, communicates, electronic and automatic control conference (IMCEC), pp 1800–1803
6. Gupta R, Bera JN, Mitra M (2010) Development of an embedded system and MATLAB-based GUI for online acquisition and analysis of ECG signal. Elsevier J Int Measur Confederation 43:1119–1126 (2010)
7. Dishongh TJ, McGrath M (2010) WSN technologies: microcontrollers, Wireless sensor networks for healthcare applications. Artech House, Boston, pp 16–34
8. Sridhara SR. Ultra-low power microcontrollers for portable, wearable, and implantable medical electronics. Texas Instruments, Inc. https://pdfs.semanticscholar.org/6da8/7a8faa359a61a68e856c347b8f52ac143a89.pdf
9. Freitas E, Azevedo A (2016) Wireless biomedical sensors networks: the technology. In: Proceedings of the 2nd world congress on electrical engineering and computer systems and science (EECSS’16), Budapest, Hungary, pp 134-1–134-8
10. Official IEEE 802.11 working group project timelines. 2016-03-23. Accessed from the original on 2016-04-07
11. Karagiannis V, Chatzimisios P, Vazquez-Gallego F, Alonso-Zarate J (2015) A survey on application layer protocols for the internet of thing. In: Transaction on IoT and cloud computing 2015, pp 1–10

Template-Based Video Search Engine

Sheena Gupta and R. K. Kulkarni

Abstract The exponential increase in video-based information has made it challenging for users to search specific video from a huge database. In this paper, template-based video search engine is proposed to improve the retrieval efficiency and accuracy of search engines. To begin with, the system splits the video sequence into eight key frames and then the fused image is created. The visual features like color and texture are extracted from the fused image and stored as complete feature set in a database. Now, the query clip is selected from the query database and then the template image is selected from the fused query image. The template query image features are compared with stored feature database using various similarity measures. The relevant retrieval experiments show that template-based video search engine using wavelet-based feature extraction gives better result in terms of average precision and recall using Euclidean distance as a similarity measure.

1 Introduction

The search for video content over the web still seems to be extremely difficult even after the success of different web search engines [1]. Generally, search engines index the metadata of videos and search them by text. However, in most cases, text-based search output is not relevant in case as videos are a spatiotemporal entity [2]. Based on the research, fused-image-based video retrieval system is an effective and efficient retrieval technique with minimum retrieval time. The fused image is a single image representing the characteristics of all key frames, and hence reduces the computational complexity of the system. However, it shows good results in terms of average

S. Gupta · R. K. Kulkarni (✉)
Department of Electronics & Communication Engineering, VESIT, Chembur, Mumbai, Maharashtra, India
e-mail: ramesh.kulkarni@ves.ac.in

S. Gupta
e-mail: sheena.gupta@ves.ac.in

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_27

precision and recall when the key frames of query are highly correlated but when they are not correlated then the results are not so satisfactory [3].

This paper presents an effective template-based video search engine which rectifies the above limitation. Fused image generation and template image selection are the two important representation schemes of query video clip used in the proposed system. So, by selecting the template image from fused query image, the retrieval search is more focused, and hence it increases the efficiency of fused-image-based video search engine. The proposed method gives higher precision and recall values for complex query video clips and the retrieval efficiency of the proposed system depends upon the template image selected.

The paper is structured as follows: The proposed system block diagram is given in Sect. 2. The framework of template-based video search engine is presented in Sect. 3. The simulation results and performance evaluation is given in Sect. 4. Finally, the conclusion and future scope are illustrated in Sect. 5.

2 Template-Based Video Search Engine

2.1 Registration Phase

In registration phase, the sample video is selected from stored video database. Then, the key frames are extracted followed by fused image creation from the sample video. The features like color moments and texture are extracted from that fused image and stored as a complete feature vector of the respective videos. This procedure is repeated for all videos stored in a video database and then the complete feature vector database is built.

2.2 Query Execution Phase

On the query side, the user gives the query video clip as input and eight key frames are extracted and fused to form a query image. From that fused query image, template image is selected and then the visual features like color moments and texture are extracted and stored as a complete feature set of that respective query video. Now, this query feature set is compared with complete feature set stored in a database using different similarity measures. Finally, the top five videos are retrieved from the videos collection. The selection of template image from fused image which contains multiple key frames in a single image that plays an important role in retrieval efficiency of system. The block diagram of the proposed template-based video search engine is shown in Fig. 1.

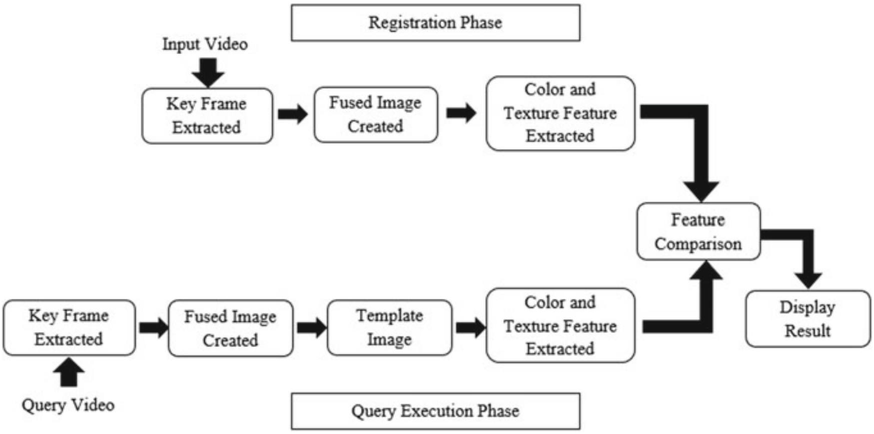


Fig. 1 Proposed framework of template-based video search engine

3 Framework of Proposed Methodology

Feature-based approach for template matching is well suited when both reference and template images had more correspondence with respect to features and control points [4, 5]. The process is explained in two parts.

3.1 Feature Extraction Using Wavelets and Color Moments

Wavelets provide multiresolution capability, good energy compaction, and adaptability to human visual characteristics and are now being adopted for various applications like industrial supervision of gearwheel, speech recognition, computer graphics, and multifractal analysis [6, 7]. Haar transforms are used to extract features as they give high energy compaction in transformed domain. The combination of energy and standard deviation is stored as a feature vector of that respective template image.

The Haar wavelet's mother wavelet function $\varphi(t)$ can be described as [8] follows:

$$\varphi(t) = \begin{cases} 1, & 0 \le t \le \frac{1}{2} \\ -1, & \frac{1}{2} \le t \le 1 \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

And, its scaling function $\varphi(t)$ can be described as

$$\varphi(t) = \begin{cases} 1, & 0 \le t \le 1 \\ 0, & \text{otherwise} \end{cases} \tag{2}$$

Color Feature. Color moment is one of the popular techniques widely used by researchers. It basically describes the color distribution of an image through its moment [9]. The three central moments mean, standard deviation, and skewness are considered for representing the feature vector [10].

3.2 Feature Extraction Using Gabor Filter and Color Moments

Gabor filters are a group of wavelets, with each wavelet capturing energy at a specific frequency and a specific direction. The scale (frequency) and orientation tunable property of Gabor filter make it especially useful for texture analysis [11]. The filters of a Gabor filter bank are designed to detect different frequencies and orientations. From each filtered image, Gabor features can be calculated and used to retrieve images. For a given image $I(x, y)$, the discrete Gabor wavelet transform is given by a convolution [11] as

$$W_{mn} = \sum_{x1} \sum_{y1} I(x1, y1) gmn * (x - x1, y - y1) \tag{3}$$

where * indicates complex conjugate and m, n specify the scale and orientations of wavelet, respectively. A feature vector f (texture representation) is created using mn as the feature components. M scales and N orientations are used and the feature vector is given in Eq. 4

$$f = \left[\sigma_{00}, \sigma_{01}, \sigma_{02} \cdots \sigma_{(M-1)(N-1)}\right] \tag{4}$$

4 Simulation Results and Performance Evaluation

4.1 Experimental Results

Template-based video search engine has been implemented using MATLAB R2017a and the performance of the proposed system is analyzed using evaluation metrics including precision, recall, retrieval efficiency, and retrieval time. The experiments are performed on a dataset of 40 videos. The collected videos contain the categories of mountain, street, ocean, and plane.

Table 1 Performance metrics for different proposed methods using Euclidean distance

Euclidean distance					
Proposed methods	Query category	Avg. precision (%)	Avg. recall (%)	Avg. retrieval efficiency (%)	Avg. retrieval time (s)
Method P_1	*Q1*	*68*	*34*	*100*	*49.03*
	Q2	44	22	100	51.83
	Q3	*66*	*33*	*90*	*56.96*
	Q4	44	22	100	49.26
Method P_2	Q1	52	26	70	52.31
	Q2	*58*	*34*	*100*	*48.50*
	Q3	46	23	90	48.26
	Q4	*64*	*32*	*90*	*46.50*
Method P_3	Q1	40	20	80	55.18
	Q2	64	32	100	64.86
	Q3	44	22	90	61.99
	Q4	56	28	80	58.84

Method P_1 is fused-image-based video search engine
Method P_2 is template-based video search engine using wavelet-based feature extraction
Method P_3 is template-based video search engine using Gabor filter feature extraction

Fig. 2 Average precision for different query categories using Euclidean distance

Figure 6a, b shows one query video clip and fused image of the query video and the template image with top five video retrieval results are shown in Figs. 7 and 8. In Tables 1, 2, and 3, the average precision, recall, retrieval efficiency, and retrieval time is calculated using three similarity measures. Figures 2, 3, and 4 show the graphical representation of average precision for all query categories (Fig. 5).

From Tables 1, 2, and 3, it is clear that the percentage of average precision and recall is higher in Proposed Method P_2 (template-based video search engine using wavelet-based feature extraction) for different query categories with Euclidean distance as a similarity measure. It is also clear that the Proposed Method P_2 gives highest precision for street (Q2) and plane (Q4) categories where the key frames are not correlated with each other, and hence the results are better in comparison to Proposed Method P_1 (fused-image-based video search engine). By selecting template image from a fused image which contains multiple key frames in a single image, the search is more focused about that particular object. The retrieval time of Proposed Method P_2 is also lesser because wavelet transform gives higher energy compaction which results in reduction in feature vector size and also reduces the retrieval time.

Table 2 Performance metrics for different proposed methods using correlation distance

Correlation distance					
Proposed methods	Query category	Avg. precision (%)	Avg. recall (%)	Avg. retrieval efficiency (%)	Avg. retrieval time (s)
Method P_1	Q1	36	18	70	51.61
	Q2	*50*	*25*	*100*	*51.34*
	Q3	*58*	*29*	*90*	*53.64*
	Q4	50	25	100	57.91
Method P_2	Q1	58	29	60	48.75
	Q2	*48*	*24*	*90*	*52.42*
	Q3	28	14	60	44.91
	Q4	*52*	*26*	*70*	*54.52*
Method P_3	Q1	42	21	60	53.20
	Q2	46	23	100	60.77
	Q3	42	21	90	53.64
	Q4	50	25	90	60.51

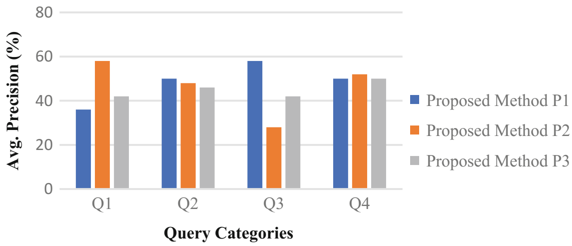


Fig. 3 Average precision for different query categories using correlation distance

Table 3 Performance metrics for different proposed methods using Minkowski distance

Minkowski distance					
Proposed methods	Query category	Avg. precision (%)	Avg. recall (%)	Avg. retrieval efficiency (%)	Avg. retrieval time (s)
Method P_1	*Q1*	*66*	*33*	*100*	*51.61*
	Q2	52	26	100	51.34
	Q3	*56*	*28*	*80*	*53.64*
	Q4	44	22	100	57.91
Method P_2	Q1	52	26	70	49.87
	Q2	*64*	*32*	*100*	*52.69*
	Q3	38	19	70	49.05
	Q4	*60*	*30*	*100*	*54.02*
Method P_3	Q1	42	21	80	58.83
	Q2	64	32	100	61.29
	Q3	42	21	90	69.66
	Q4	52	26	70	56.17

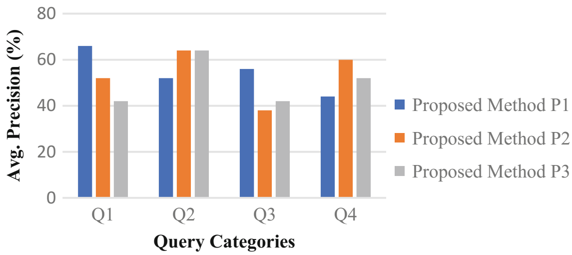


Fig. 4 Average precision for different query categories using Minkowski distance

It is observed from Fig. 5 that Proposed Method P_3 (template-based video search engine using Gabor-filter-based feature extraction) gives higher average retrieval efficiency in comparison to Proposed Method P_1 but the retrieval time is more because Gabor wavelet transform requires large space for storage, and hence increases the computational time.

4.2 Performance Evaluation

The proposed method is analyzed using three similarity measures, viz., Euclidean, correlation, and Minkowski distance.

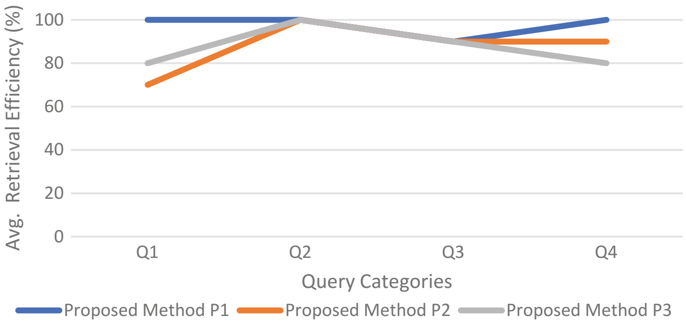


Fig. 5 Average retrieval efficiency for different query categories using Euclidean distance

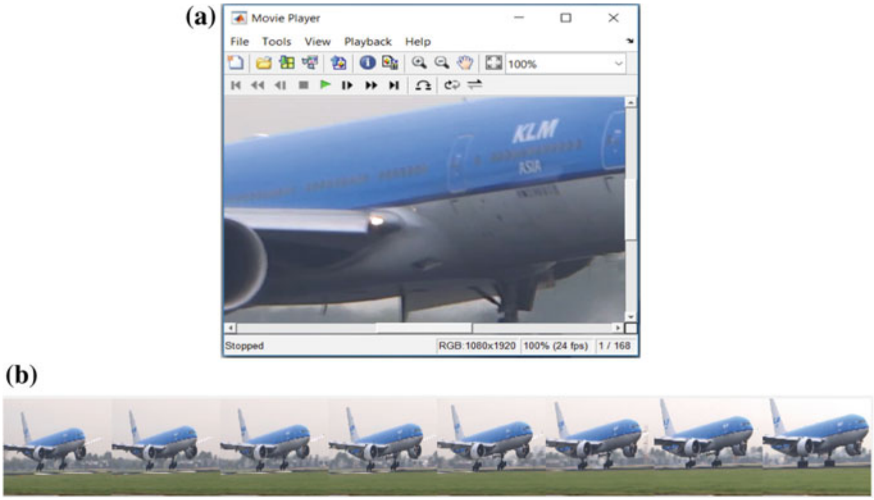


Fig. 6 **a** Proposed framework for template-based video search engine. **b** Key frames extraction for the query input video clip

Fig. 7 Shows template image for feature extraction

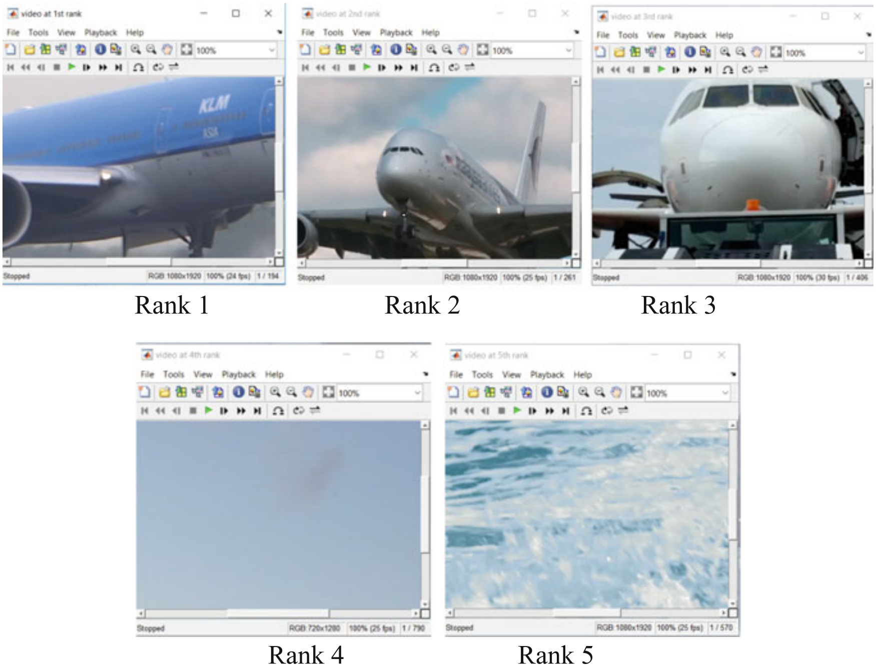


Fig. 8 Shows the top five ranked videos of the query video shown in Fig. 6a

5 Conclusion and Future Scope

Different methods of template-based video search engine have been tested and the average precision, recall, retrieval efficiency, and retrieval time are analyzed. Best precision is given by template-based video search engine using wavelet-based feature extraction (Proposed Method P_2) with Euclidean distance as similarity measures along with less retrieval time. So, by choosing template image from fused query image, the retrieval search is more focused and accurate. The video search engine using template matching is also compared with fused-image-based video search engine and result shows there is improvement in precision values for complex query video clips.

In future, the accuracy of the proposed system can be increased further by increasing number of key frames and by extracting more visual features.

References

1. Shanmugam TN, Rajendran P (2009) An enhanced content-based video retrieval system based on query clip. Int J Res Rev Appl Sci 1(3). ISSN:2076-734X, EISSN:2076-7366
2. Hu W, Xie N, Li L, Zeng X, Maybank S (2011) A survey on visual content-based video indexing and retrieval. IEEE Trans Syst Man Cybern Part C Appl Rev 41(6):797–819
3. Gupta S, Kulkarni RK (2017) Fused image based video search engine. Int J Video Image Process Netw Secur (IJVIPNS-IJENS) 17(5) 174305-2929 (October)
4. Mahalakshmi T, Muthaiah R, Swaminathan P (2012) Review Article: An overview of template matching technique in image processing. Res J Appl Sci Eng Technol 4:5469–5473. ISSN: 2040-7467
5. Nazil P, Kumar D, Bhardwaj I (2013) An overview on template matching methodologies and its applications. Int J Res Comput Commun Technol 2(10):988–995
6. Thepade SD, Yadav N (2015) Novel efficient content based video retrieval method using cosine-haar hybrid wavelet transform with energy compaction. In: International conference on computing communication control and automation. IEEE. 978-1-4799-6892
7. Deepa T, Girisha H (2014) Image compression using Hybrid wavelet Transform and their Performance Comparison. Int J Mod Eng Res (IJMER) 4(6):6–12. ISSN: 2249–6645
8. Kekre HB, Thepade SD, Gupta S (2013) Content based video retrieval in transformed domain using fractional coefficients. Int J Image Process (IJIP) 7(3):238–274
9. Padmakala S, Anandha Mala GS, Shalini M (2011) An effective content based video retrieval utilizing texture, color and optimal key frame features. In: International conference on image information processing (ICIIP 2011). IEEE. 978-1-61284-861
10. Geetha P, Narayanan V (2011) An effective video search re-ranking for content based video retrieval. IEEE. 978-1-4673-0131
11. Ansari A, Mohammed MH (2015) Content based video retrieval system-methods, techniques, trends and challenges. Int J Comput Appl 112(7) (Feb)

Gray-Level Feature Based Approach for Correspondence Matching and Elimination of False Matches

R. Akshaya and Hema P. Menon

Abstract Matching of interest points (feature points) is a basic and very essential step for many image processing applications. Depending on the accuracy of the matches, the quality of the final application is decided. There are various methods proposed to tackle the problem of correspondence matching. In this paper, a method that makes use of the textural features, mainly gray-level features with respect to a pixel's neighborhood has been discussed for point matching with an emphasis on its use for image registration. Feature points are obtained from the images under consideration using SURF and are matched using gray-level features. Then, the false matches are removed using a graph-based approach.

1 Introduction

With a lot of images that are captured nowadays, to analyze them or to make a meaningful conclusion from it, these images should be aligned first. This process of aligning the images with one as a reference image is called Registration [1]. During Registration, all the images are mapped to a common coordinate system. The images can be different due to various reasons like different viewpoints, captured at different depths, taken at different timings, or due to the changes in the capturing modes [2]. Generally, Registration is divided into two types called Rigid and Non-rigid registration. Rigid registration is used in the case when the is no change in the objects size and shape and Non-Rigid registration is used in case if there is a change in the size and shape of the object. Registration is used in various applications like Medical images, fusion, Super-resolution, shape analysis, biometrics, 3D reconstruction, etc., [3–7].

R. Akshaya · H. P. Menon (✉)
Department of Computer Science and Engineering, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Coimbatore, India
e-mail: p_hema@cb.amrita.edu

R. Akshaya
e-mail: cb.en.p2cvi16001@cb.students.amrita.edu

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_28

In simple words, Registration involves identifying the points that are the same from both the images, define a transformation function to align the target image with respect to the source image. The image that is used as a reference image is called the source image, it is with respect to this image, all other images are aligned. The other image that is aligned with respect to the source image is called the target image. The first step in Registration is very important, that is finding the points that are the same or the corresponding points in the images. This is the basic step that decides the success of Registration. Correspondences in the image are found using feature descriptors. This paper mainly discusses how to match the points and remove the points that are falsely matched [8–10].

2 Related Work

Finding out corresponding points are almost similar to a search problem, if a point is there in the first image, a match for it is searched in the second. There are different methods that have been proposed to find the match, for example, pixelwise search, which is computationally huge, feature-based methods where features are extracted and match these to find the corresponding points or graph methods to do the same as the structure of the object remains the same in both the images. The matching can also be done using the neighborhood of the pixels, like the template matching or the using textural features of the neighborhood. This could possibly not be done on every point as first it is computationally complex and second the pixels in the neighborhood will mostly remain the same approximately. So, feature points that are unique to the image are extracted. SURF is one such method. Speeded-up robust features is a local feature descriptor. It is a much faster version of the SIFT extractor. The DOG is used in case of SIFT and LOG with a box filter is used in case of SURF. The reason why SURF is faster is that convolution of a box filter is as simple as calculating the integral images and can be done simultaneously at all scales [11]. The textural features from a neighborhood can be extracted using the Gray-Level Co-occurrence Matrix. The GLCM is calculating features using the statistical methods. It essentially uses the spatial relation between the pairs of pixels at different orientations to define a matrix. For example, if the pixel pair that is next to each other are considered, then the angle between them is 0°, and likewise for different angles. From this GLCM, various features like contrast, energy, homogeneity, etc., are extracted [12].

There are a lot of papers that discuss how to match the points in the given image. Patricio et al. proposed a method where matching between the pair of images, which was done for stereo vision matching. The matching between the images was done based on the correlation computed between the point and the neighborhood. The pixels that were not similar to the neighborhood were avoided for computation, a method of the adaptive neighborhood is defined, which is based on the similarity of the pixel with its neighborhood. The size and shape of the neighbourhood are decided based on the content it has. The algorithm is compared with algorithms like the SAD, HIR, SMW, has reduced error rate [13]. Silva et al. have proposed a method to find out the corresponding interest points based on the Morse complex. It uses a topological operator called the Local Morse Context (LMC) to mainly explore the structural information. It mainly reduces the incorrect matches by giving a confidence number. Here, the interest points are defined by a Morse cell adding more importance to the neighborhood pixels using subcomplex of a complex, where standard graph solution is used. Eigen decomposition is used to solve the problem of inexact graph matching to handle noises and inconsistencies [14].

3 Implementation

3.1 Proposed Method

In the proposed method, the corresponding matches in both images were found using the points extracted using descriptors. The general flow diagram of the method is shown in Fig. 1.

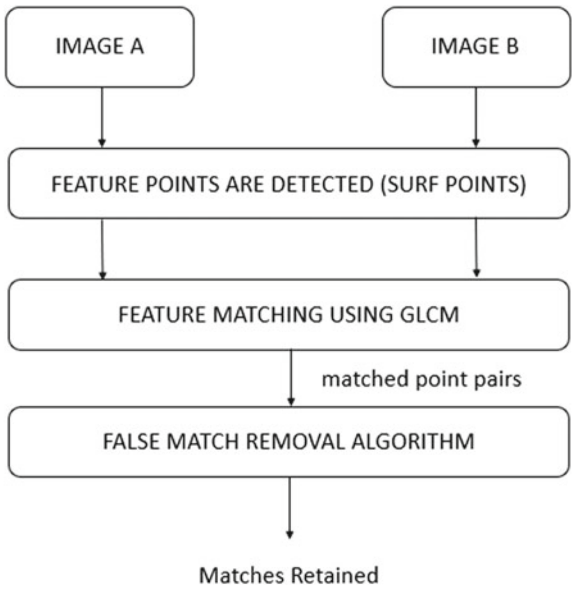


Fig. 1 Above are the general steps in the proposed algorithm

3.2 Dataset Details

The datasets contain a set of 6 ppm images of 640 × 800 resolution. These images are taken at different Viewpoints. And the other data is also a set of 6 ppm images of size 1000 × 7000 [15].

3.3 Methodology

The descriptor that is used to extract the point here is SURF. SURF feature descriptor is applied to both the images and a set of unique points is got. Say we have p_1 points from image A and p_2 points from image B. p_1 and p_2 need not be essentially the same. So, we select the image with the lesser number of points from images A and B. Since we are finding a match for all the points that are extracted, using the image with more number of points will lead to mapping a point to more of number of points from which arriving at the correct match might be difficult.

For all the points that are extracted using the Surf features, GLCM is calculated on a neighborhood of 11 * 11 and an angle of 0°. Once the GLCM is obtained, features like the correlation, variance, energy, contrast, mean, etc., are calculated for it. The gray features of every point in the image with a lesser number of features points is compared with the features of every other point on the other image. The comparison is done using the Euclidean distance between the features. The point is then matched with the point on the other image which has got the minimum difference when compared to all other points. This procedure is then done for all the points present, the point that is matched with the previous point will be still available for the points that come after that due to the fact that the previous match could be a false one. So, now we have p matches, [min ($p1$, $p2$)].

All the matches need not be essentially the true matches. So, we apply an algorithm proposed by Zakharov et al. [16] for the removal of false matches. A matrix of distance is calculated, G. And singular value decomposition is done on this distance matrix. The eigenvalue matrix is then replaced with an identity matrix to get a matrix P. A matrix of correspondence, K is also created where matrix has the value equal to one when the corresponding indices of the points match, for example when point with the index 1 in the first set of points match with the point having index 2 in the second set the (1, 2) position of the matrix is equal to one. Then a dot product is applied on the matrices P and K. If the value crosses a particular threshold then the points at the corresponding indices are said to be true matches. The threshold set in this experiment is 0.5, so after performing a dot product if the value crosses 0.5 then its considered as corresponding points. The above algorithm is applied to every 5% of the points to get the results. Thus, the points corresponding points and the false matches are identified.

4 Experimental Results

The experiments were performed on a set of images from the dataset. The output images are shown below. Figure 2 shows the images that show first 50 matched points using GLCM features without the removal of false matches. Figure 3 shows the output of the first 50 points matched after false match removal. Figure 4 shows the final output of the image 1 and 2 in the dataset [15]. Figure 5 is the final output of the images 1 and 2 in the dataset (Fig. 6).

Table 1 shows the number of points matched using the GLCM algorithm and points retained after false match removal in the images of dataset [15].

Fig. 2 Matched points (50) got after matching from GLCM

Fig. 3 Output of matching first 50 points of images

Table 1 Showing the number of pairs extracted and the number remaining after

Sample image pairs considered	Number of matched points	Number of matched retained after false match removal
Pair 1	1463	240
Pair 2	1463	158
Pair 3	1463	10
Pair 4	1463	7

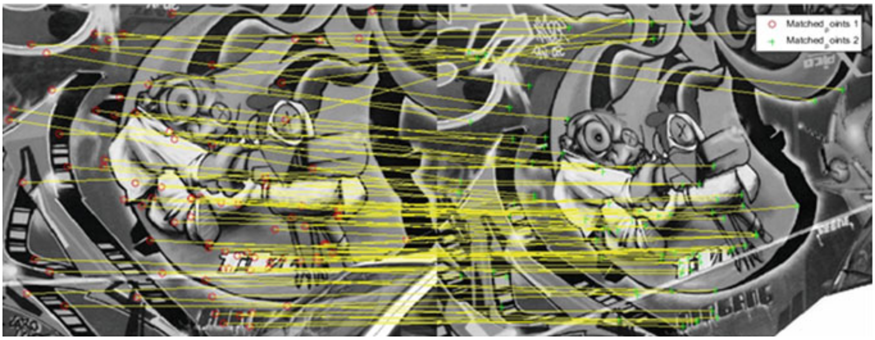


Fig. 4 Output for corresponding points matching in images

Fig. 5 Output for corresponding points matching of images

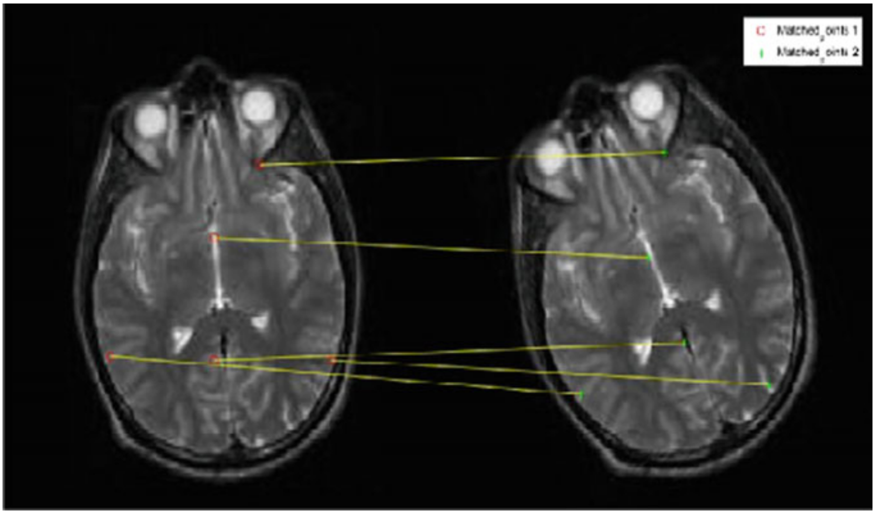


Fig. 6 Output for corresponding points matching of images (BRAIN)

3	3
1	3

2	1
1	6

3	1
1	4

Fig. 7 Confusion matrix for the brain, painting and bikes dataset, respectively

4.1 Analysis

Table 1 shows the points that were retained by the algorithm as true matches. To evaluate the performance of the algorithm, further analysis has to be carried out. There might be matches that are true and removed or vice versa, so it is necessary to analyze the results using True positives, True negatives, False positive and False negatives. Analyzing the above results visually is difficult as it is very clustery. Hence, a subset of points is used to make the analysis easy. Figure 7 given is the confusion matrix of different datasets.

4.2 Inference

Experiments have been performed with the image 1 in the dataset as the reference image and the matched points remain a constant because the minimum number of points from the two images are considered and mostly image 1 had the minimum number of points. The observation made from the results are, the false matches that were present in the image after matching were removed on the application of the false match removal algorithm. And the number of points that were obtained after the false match removal reduced as the viewing angle deviated a lot with respect to the reference image. This can be seen in the above-tabulated results too. The viewing angle increases from image 2 to 5. From the confusion matrix, it can be seen that most of the false matches are removed. The above algorithm works better in case of textural images, whereas in the case of medical images brain MRI the algorithm also removes most of the true matches. Thus, the points obtained can be given to any registration algorithm to register the images.

5 Conclusion

This paper presents a method to match the interest points from the images and remove the false matches from them using the textural features from the images for a particular neighborhood. The features extracted help us to define the area around the pixel of interest and matches the pixel with the same sort of features in the other image. These matched feature points are given to an algorithm that removes the false matches present in the image. As a future work, the false matches that are still existing will be eliminated to make the end result much better.

References

1. Zitova B, Flusser J (2003) Image registration methods: a survey. Image Vis Comput 21(11):977–1000
2. Brown LG (1992) A survey of image registration techniques. ACM Comput Surv (CSUR). 24(4):325–376
3. Ong EP, Xu Y, Wong DW, Liu J (2015) Retina verification using a combined points and edges approach. In: 2015 IEEE international conference on image processing (ICIP), 27 Sept 2015, pp 2720–2724. IEEE
4. Menon HP, Narayanankutty KA (2015) Comparative performance of different perceptual contrast fusion techniques using MLS. Int J Biomed Eng Technol 18(1):52–71
5. Shwetha R, Rajathilagam B (2015) Super resolution of mammograms for breast cancer detection. Int J Appl Eng Res 10(1):21453–21465
6. Huang X, Zhang J, Fan L, Wu Q, Yuan C (2017) A systematic approach for cross-source point cloud registration by preserving macro and micro structures. IEEE Trans Image Process 26:3261–3276
7. Arathi T, Parameswaran L (2014) Image reconstruction from 2D stack of MRI/CT to 3D using shapelets. Int J Eng Technol (IJET). 6(1):2595–2603
8. Jain V, Li X (2004) Point matching methods: survey and comparison. Project report for CMPT 8888
9. Menon HP, Nitheesh AS (2017) Structural matching of control points using VDLA approach for MLS based registration of brain MRI/CT images and image graph construction using minimum radial distance. In: The international symposium on intelligent systems technologies and applications. Springer, Cham, pp 356–369
10. Menon HP (2017) An analysis on the influence that the position and number of control points have on MLS registration of medical images. In: International symposium on signal processing and intelligent recognition systems. Springer, Cham, pp 47–56
11. Bay H, Tuytelaars T, Van Gool L (2006) Surf: speeded up robust features. In: European conference on computer vision. Springer, Berlin, pp 404–417
12. Mohanaiah P, Sathyanarayana P, GuruKumar L (2013) Image texture feature extraction using GLCM approach. Int J Sci Res Publ 3(5):1
13. Patricio MP, Cabestaing F, Colot O, Bonnet P (2004) A similarity-based adaptive neighborhood method for correlation-based stereo matching. In: 2004 international conference on image processing, 2004. ICIP'04, vol 2. IEEE, pp 1341–1344
14. da Silva RD, Schwartz WR, Pedrini H, Pulido J, Hamann B (2015) A topology-based approach to computing neighborhood-of-interest points using the Morse complex. J Vis Commun Image Represent 30:299–311
15. Image Details. http://www.robots.ox.ac.uk/~vgg/data/data-aff.html
16. Zakharov AA, Tuzhilkin AY, Zhiznyakov AL (2015) Finding correspondences between images using descriptors and graphs. Procedia Eng 1(129):391–396

A New Approach for Image Compression Using Efficient Coding Technique and BPN for Medical Images

M. Rajasekhar Reddy, M. Akkshya Deepika, D. Anusha, J. Iswariya and K. S. Ravichandran

Abstract Medical images produce a digital form of human body pictures. Most of the medical images contain large volumes of image data that is not used for further analysis. There exists a need to compress these images for storage issues and to produce a high-quality image. This paper discusses an image compression using Back Propagation Neural network (BPN) and an efficient coding technique for MRI images. Image compression using BPN produces an image without degrading its quality and it requires less encoding time. An efficient coding technique—Arithmetic coding is used to produce an image with better compression ratio and redundancy is much reduced. Back-Propagation Neural Network with arithmetic coding gives the better results.

1 Introduction

Image compression is a compression technique which reduces storage space of an image without degrading the quality of the on diagnosis of diseases, hence it is important to maintain the image quality [1]. An efficient compression technique is required for the medical image to produce a high-quality image [2]. The MRI images are very large in size so it is necessary to reduce the cost of storage to increase the transmission speed [3].

The Back-Propagation Neural Network (BPN) is the most widely used neural network for image compression [4]. An image is partitioned into blocks without overlapping and fed into neural network [5]. The blocks are used to train the neural network and gives the better reconstructed image [6].

Arithmetic coding is a form of entropy encoding used in all type of data compression [7]. Arithmetic coding encodes the frequently seen symbols with fewer bits [8]. Flexibility is an advantage of arithmetic coding [9].

M. Rajasekhar Reddy (✉) · M. Akkshya Deepika · D. Anusha · J. Iswariya · K. S. Ravichandran
School of Computing, SASTRA Deemed University, Thanjavur, India
e-mail: rajasekharmanyam04@gmail.com

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_29

In this paper, an image is compressed using BPN and arithmetic coding. An image is compressed using BPN and further compression is done by using arithmetic coding [10]. The final image gives the better Peak Signal-to-Noise Ratio (PSNR).

2 Existing Methods

2.1 Back-Propagation Neural Network

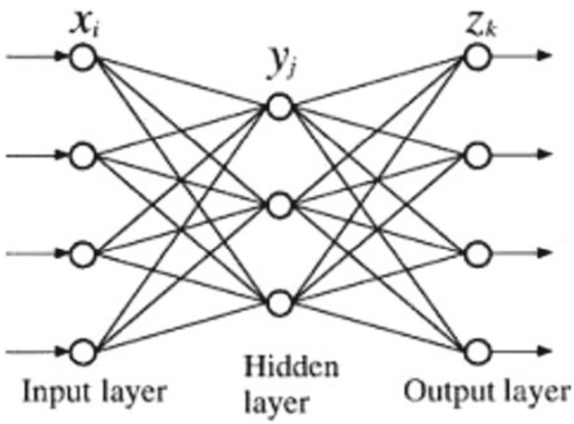
There are three layers in neural network. The size of input and output layers are same [11]. The hidden layer is fully interconnected with input and output layer [12]. Image encoding is done for each pixel of the block which is transmitted to the receiving side where reconstruction is done [13]. Compression is achieved at the hidden layer and decompression is achieved in output layer [14]. The two-activation function used in neural network is "tensing" for hidden layer and "purelin" for output layer (Fig. 1).

The neural network is initialized and trained by using the learning rate and goal with a set of desired values [5]. The default performance measures are mean square error. There will be a decrease in performance of the neural network when number of weights increases [15]. Neural network-based image compression reduces the search space and encoding further improve computation time [16].

2.2 Arithmetic Coding

Arithmetic coding is used in all types of data compression algorithms. It has some advantages over Huffman coding where it gives better compression ratio and redundancy is much reduced [17]. Unlike Huffman coding, discrete number of bits for each

Fig. 1 Back-Propagation Neural Network



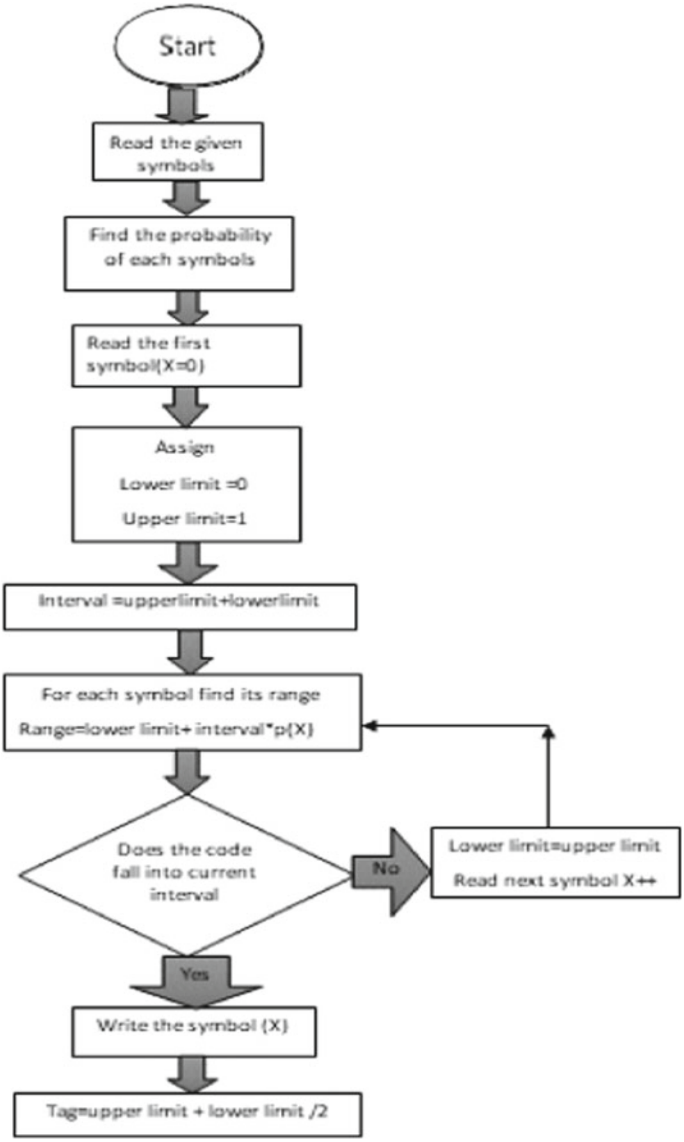


Fig. 2 Arithmetic coding

code is not used in arithmetic coding [8]. Each symbol is assigned with an interval, and the probability of each symbol is calculated [18]. Starting with lower limit 0 and upper limit 1, each interval is divided into several subintervals [7] (Fig. 2).

```
BEGIN
Lower limit = 0; upper limit = 1; range = 1;
While (symbol! = terminator)
{
Get (symbol);
Lower limit = lower limit + range * Range_lowerlimit(symbol);
High = lower limit + range * Range_upperlimit (symbol);
Range = upper limit − lower limit;
}
Output code word
Tag = lower limit + upperlimit/2;
END. [9]
```

3 Proposed Method

Image compressing is done using BPN and arithmetic which gives better results.

3.1 Procedure

1. Read an input medical image.
2. Resize it to Standard Image size and normalize the image.
3. The image is partitioned into nonoverlapping block and it is fed into neural network as input.
4. Initialize the weights and targets.
5. Set the training parameters.
6. Train the neural network.
7. Apply arithmetic coding to the resultant image.
8. Save the arithmetic codes.
9. Convert these codes using inverse arithmetic coding.
10. Obtain the reconstructed image.
11. Compute the compression ratio.
12. Calculate the PSNR value.
13. Note the execution time (Fig. 3).

4 Performance Measure

The parameters used to measure performance are Mean Square Error (MSE) and Peak Signal-to-Noise Ratio (PSNR).

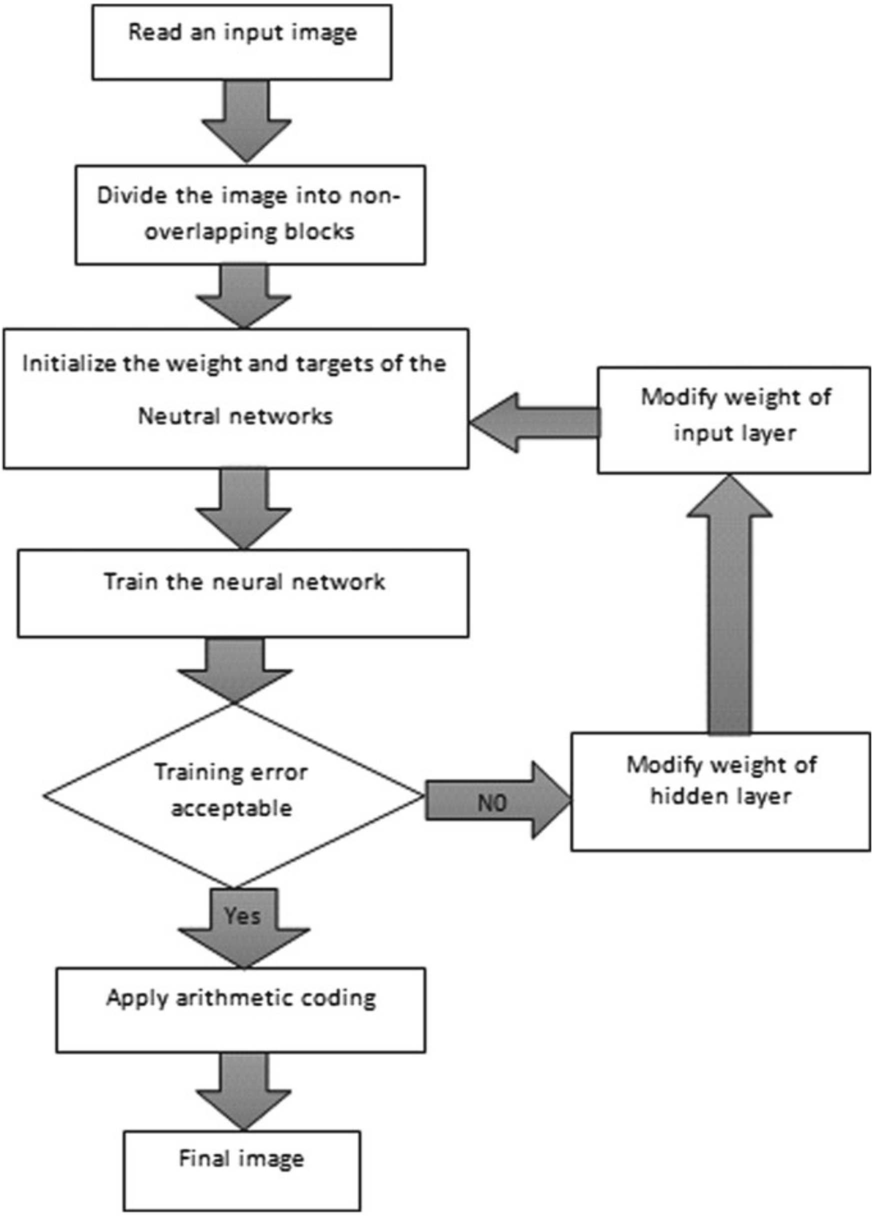


Fig. 3 Back-Propagations Neural Network and arithmetic coding

$$\mathrm{MSE} = 1/MN \sum_{y1=1}^{m} \sum_{x1=1}^{n} \left[I(x1, y1) - I'(x2, y2) \right]^2,$$

where $M \times N$ is an image.

PSNR is defined as

$$\text{PSNR} = 20\ \log_{10}\left[255/\text{sqrt}(\text{MSE})\right]$$

Quality of a reconstructed image is commonly measured by PSNR value [19].

5 Results

See Table 1, Figs. 4 and 5.

Table 1 PSNR values and execution time for existing and proposed techniques

Input image	PSNR using BPN (dB)	PSNR using BPN and arithmetic (dB)	ET using BPN (s)	ET using BPN and arithmetic (s)
Image (1)	29.906	52.41	108.633	121.68
Image (2)	30.41	53.70	79.246	118.98
Image (3)	30.42	52.74	105.216	124.189
Image (4)	30.44	54.11	79.894	129.17
Image (5)	30.41	53.65	127.434	112.00
Image (6)	31.05	54.34	117.893	130.88
Image (7)	31.46	54.99	113.16	97.40
Image (8)	31.63	54.88	76.396	130.65

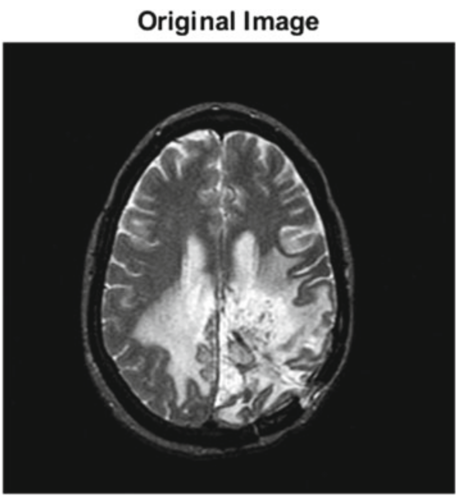


Fig. 4 Input image (5)

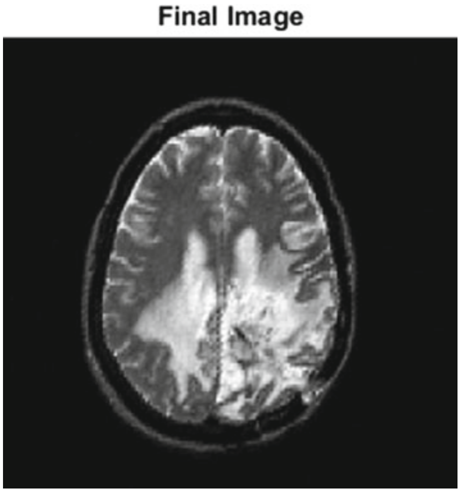


Fig. 5 Reconstructed image (5)

6 Conclusions

A new approach for image compression using BPN and arithmetic coding for medical images produces a better quality image. The PSNR value of the proposed method is increased as compared with an existing one. Thus, the hybrid image compressing using BPN and arithmetic coding gives better results than image compressing using BPN alone.

References

1. Dougherty G (2009) Digital image processing for medical applications, 1st edn. Cambridge University Press, Cambridge, pp 9–10
2. Shapiro JM (1993) Embedded image coding using zero trees of wavelet coefficients. IEEE Trans Sig Process 41(12):3445–3462
3. Maha Lakshmi GV, Rama Mohana Rao S (2012) A novel algorithm for image compression based on fractal. Image Compression EJSR 85(4):486–499. ISSN: 1450–216X, Eurojournals
4. Savkovic-Stevanovic J (1994) Neural networks for process analysis and optimization: modeling and application. Comput Chem Eng 18(11–12):1149–1155 (14 Ref.)
5. Soliman HS, Omari M (2006) A neural networks approach to image data compression. Appl Soft Comput 6(3):258–271
6. Maha Lakshmi GV, Rama Mohana Rao S (2013) A novel algorithm for image compression based on fractal and neural networks. Int J Eng Innov Technol (IJEIT) 3(4):8–15
7. Kavitha V, Easwarakumar KS (2008) Enhancing privacy in arithmetic coding. ICGST-AIML J 8(I):23–28
8. Masmoudi A, Masmoudi A (2013) A new arithmetic coding model for a block-based lossless image compression based on exploiting inter-block correlation. Sig Image Video Process 1–7
9. Witten IH, Neal R, Gleary JG (1987) Arithmetic coding for data compression. Commun ACM 30(4):520–540

10. Vilovic I (2006) An experience in image compression using neural networks. In: Proceedings of the 48th international symposium ELMAR-2006 focused on multimedia signal processing and communications, June 2006. IEEE Press, Zadar, Croatia, pp 95–98
11. Meyer-Bäse A, Jancke K, Wismüller A, Foo S, Martinetz T (2005) Medical image compression using topology-preserving neural networks. Eng Appl Artif Intell 18(4):383–392
12. Carrato S (1992) Neural networks for image compression. In: Neural networks: advancement and application, 2 edn. Galenbe Pub, North-Holland, Amsterdam, pp 177–198
13. Srikala P, Umar S (2012) Neural network based image compression with lifting scheme and RLC. Int J Res Eng Technol 1(1):13–19
14. Osowski S, Waszczuk R, Bojarczak P (2006) Image compression using feed forward neural networks—hierarchical approach. In: Natural to artificial neural computation, vol 3497 of Lecture notes in computer science. Springer, Berlin, Germany, pp 1009–1015
15. Ashraf R, Akbar M (2007) Adaptive architecture neural nets for medical image compression. In: Proceedings of the IEEE international conference on engineering of intelligent systems (ICEIS '06), Islamabad, Pakistan, pp 1–4
16. Northan B, Dony RD (2006) Image compression with a multiresolution neural network. Can J Electr Comput Eng 31(1):49–58
17. ISO/IEC 14496-2 FPDAM 1 Information Technology. Generic Coding of Audio-Visual Objects—Part 2: Visual, July 1999
18. Rubin F (1979) Arithmetic stream coding using fixed precision registers. IEEE Trans Inf Theory 25:672–675 (Data compression conference)
19. Miaou S-G, Ke F-S, Chen S-C (2009) A lossless compression method for medical image sequences using JPEG-LS and interframe coding. IEEE Trans Inf Technol Biomed 13(5):818–821

Person Identification Using Iris Recognition: CVPR_IRIS Database

Usha R. Kamble and L. M. Waghmare

Abstract In recent days iris is as one of the useful traits for authentication using biometric recognition. Almost all publicly available iris image databases contain data correspondent to heavy imaging constraints and suitable to evaluate by various algorithms (CASIA Iris Image Database [1]; Proença et al. in IEEE Trans Pattern Anal Mach Intell 32(8):1529–1535, 2010 [2]). This paper is first to prepare our own IRIS IMAGE DATABASE and make the availability of it to the new researchers. We have thus proposed our own database named as CVPR_IRIS DATABASE with not heavy imaging constraints. This set up contains Iris Image Capture IR Sensitive CCD camera by which eye images are captured in the visible lighting conditions with IRLED source at a distance. Clear images from this database are separated from noisy images by quality assessment technique. Thus proposed CVPR_IRIS DATABASE containing total 485 eye images of 49 individuals is made available for researchers concerned with the implementation of algorithms based on research in iris recognition. Second using proposed database we have proved good accuracy using 2D Discrete Wavelet Transform.

1 Introduction

Reliable personal identification using iris as a biometric is recently an active topic in the research area of pattern recognition. It is because of increase in the demand of enhanced security of human being in daily life. For last two decades many of the researchers have worked on improving accuracy, speed, storage cost, and robustness [3–15]. Most of these algorithms are categorized into four groups [14] such as phase-based methods, zero crossings representations, texture analysis, and low-intensity variations. Bertillon put forth the main idea of person identification by using iris

U. R. Kamble (✉) · L. M. Waghmare
Shri Guru Gobind Singhji Institute of Engineering and Technology, Nanded, India
e-mail: urkamble@sggs.ac.in

L. M. Waghmare
e-mail: lmwaghmare@yahoo.com

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_30

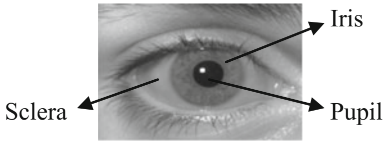


Fig. 1 Front view of human eye

recognition. After few years John Daugman [3, 6] published his rigorous work in this area. The popularity of iris as biometrics grows because of several advantages of iris over other biometrics such as palm finger prints, hand, and face, etc. Use of iris for person identification in public sectors, government and nongovernment offices, state control lines proves that iris is most accurate from other biometrics. Currently, deployed systems rely mainly on good quality images, captured in a stop-and-stare interface, at close distances and using near infrared (700–900 nm) wavelengths [2]. Our purpose is to explore new imaging process and acquire clear images. These images are to be used for the research in iris recognition. We have large number of experiments conducted on this database and reported in the literature.

Front view of the iris structure of human eye is shown in Fig. 1 consisting of iris placed in between pupil and sclera. Iris as a biometric outperforms over other biometrics because of few reasons such as it remains stable throughout one's lifetime, its uniqueness, flexibility, non-invasive, and reliable nature. The remainder of this paper is organized as follows:

Section 2 briefly describes about related work and architecture of iris recognition. Section 3 gives description of proposed work: Preparation of THE CVPR_IRIS DATABASE and database availability. Section 4 proposes algorithm using DWT reports our experiments and discusses the results. Section 5 presents the conclusions.

2 Related Work

Daugman [3, 6] the first researcher with his lot of efforts has given to the world an algorithm based on iris codes from features using 2D Gabor Wavelets. The author has used Integro differential operators to detect the center and diameter of iris. A simple Boolean Exclusive-OR operator is used as Hamming Distance for matching. For perfect match the Hamming Distance should be equal to zero. The algorithm proves accuracy of more than 99.9%. Time required for iris identification shown is found to be less than one second. The algorithm of Boles and Boashash [4] in his research work set of one-dimensional signal is encoded by using a zero crossing transformation at different resolution levels. To determine the overall dissimilarity between two iris codes, the average of the dissimilarity at each resolution level is obtained. The author reported 100% verification and identification accuracy with the experiments conducted on 11 iris images. But the source of the testing iris images was not indicated.

In paper [7] by Li Ma, Tieniu Tan, Yunhong Wang describes a new scheme for iris recognition. Here the quality of each image in the input sequence is assessed and a clear iris image is selected from a sequence if iris images. To capture local iris bank of spatial filters, whose kernels are suitable for iris recognition is then used. A comparative study of existing methods is done on an iris image database. Database includes 2255 sequences from 213 subjects and algorithm proposed highly better performance.

2.1 Iris Recognition System

The system has acquisition and preprocessing in the first module. Second module is the image verification module which has pattern matching and identification module. The second module compares input iris mages with known irises in the database. And decides if it is in the database or not. Preprocessing of iris includes correct segmentation of iris. This segmented iris is unwrapped by using certain methods. Some of the authors reported their algorithms without normalization also.

3 Description of Proposed Work: Preparation of CVPR_IRIS DATABASE

Our own iris image acquisition setup and imaging framework is as shown in Fig. 2. It consists of camera with IR LED source and CPU. The specifications of setup of the imaging framework are given in Table 1. As illustrated in Fig. 2 this framework was installed in both natural and artificial lighting sources.

We placed several marks on the floor each after 1 m up and to 10 m from acquisition device and asked for volunteers for the image acquisition processes. When planning CVPR_IRIS database we had three basic concerns to acquire images. First, all subjects are fixed at some distance. Second varying very small imaging distances, and third having normal lighting environments. All subjects are Indian, 90% subjects are of age groups of about 20 years old and 10% are from 35 to 45 years old. The

Fig. 2 Imaging framework

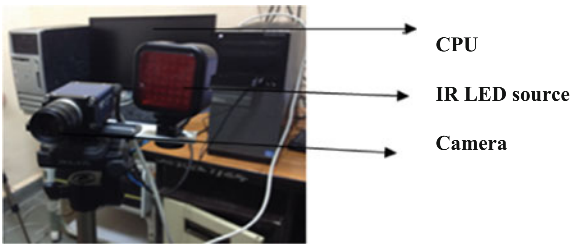
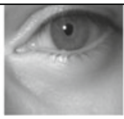
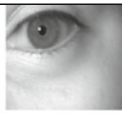


Table 1 Specifications of set up of imaging framework

Image capture in visible light with IR source Intel®	
Image type	BMP file format
Image dimension (number of pixels)	1024×768
Distance between camera and object	20 cm approximate, with variation of ± 3 cm
Distance between camera and infrared source	8 cm
Compression format	Bmp
Camera device	Onboard memory—8 MB Iris Image Capture IR Sensitive CCD camera, CCD Monochrome Camera
CPU	Intel Core, I7, Windows 10 Home, 8 GB, DDR4, 1 TB
Focal length	60 mm

Table 2 Proposed CVPR_IRIS DATABASE: 5 images of person 2 from left eye only

Example	Total Images	Wavelength	Varying Distances	Acquisition Device
				
2L1.BMP	2L2.BMP	2L3.BMP	2L4.BMP	2L5.BMP

database is constructed with 485 images from 49 individuals. For 48 individuals, we have captured five images of right and five images from left eyes. For one individual, we have captured only five images of left eye and images are numbered as 1L1.bmp to 1L5.bmp. Similarly, right and left eye images of *i*th person are numbered as iR1.bmp to iR5.bmp and iL1.bmp to iL5.bmp respectively.

Database availability: This will help others to obtain the same data sets and replicate your experiments. We suggest the following address reference for database availability: CVPR LAB, Department of Electronics and Telecommunication Engineering, SGGS Institute of Engineering & Technology, Nanded (MS), India.

Year = "2016–18",
Mail id: urkamble@sggs.ac.in

Table 2 shows five images of person 2 from left eye only naming L1 for person 1 left eye image. Likewise in proposed database R1 is for person 1 right image.

4 Experiments Using Proposed CVPR_IRIS DATABASE

While capturing iris images unfortunately we cannot have one image of good quality but we have captured sequence of images from single eye. The problem is to select more clear, good quality images. In the process of capturing images for our database we have some of the images that are defocused images, blurred images, motion blurred images, and partially occluded images because of upper eyelids. We have separated out all clear images. This data we are keeping for study and analysis. Though we have different database publicly available the related work for quality assessment of iris images is not enough and still it is limited.

Hence to assess and select image of good quality we have used fourier descriptors that is effective scheme by analyzing the frequency distribution of the iris image. Daugman [3, 6] in his work, iris images are assessed by measuring the total high-frequency power in the 2D Fourier spectrum. But he did not mention a detailed description of his method. Alexandre [2] and group detected 14 different noise factors that are classified into local or global category. Hence to assess and select image of good quality we have used fourier descriptors [12] that is effective scheme by analyzing the frequency distribution of the iris image.

4.1 Image Quality Assessment

All clear and properly focused iris images should have uniform frequency distribution. Fourier descriptor is defined by DR Eq. (1)

$$\mathrm{DR} = \left[(fl + fm + fh); \frac{fm}{fl + fh} \right] \tag{1}$$

$$f_i\ 003\mathrm{D} = \iint_{\Omega} |F(u, v)| \mathrm{d}u \mathrm{d}v \quad i = 1, 2, 3 \tag{2}$$

$$\Omega = \left\{ (u, v) \middle| f_{1i} < \sqrt{u + v^2} <= f_{2i} \right\} \tag{3}$$

$F(u, v)$ is the 2D Fourier spectrum of an iris region. fl, fm, fh are the power of low, middle and high frequency components. f_{1i} and f_{2i} are the radial frequency pairs and bound the range of the corresponding frequency components. We have selected 64×64 from right iris regions. Ideal values of d are less than $d = [6.8; 0.425]$. For larger DR values image must be discarded considering unfocussed image.

4.2 Iris Preprocessing

It consists of segmentation of iris and normalization of iris. The main objective of the segmentation step is to differentiate iris pattern from the rest of the eye image. Correctly finding the inner and outer boundaries of iris pattern is important in all iris recognition systems. We have used Daugman's Integro Differential operator for segmentation of iris.

In normalization the iris region is transformed into fixed size. We used Daugman's rubber sheet model for the normalization step.

4.3 Feature Extraction Using DWT

CVPR_IRIS database consists of image size 1024×768. For one-dimensional discrete signals 1D Discrete Wavelet Transform is mainly preferred. For images of any type two-dimensional Discrete Wavelet Transform is needed to use. 2D wavelet transform obtained by separable products of scaling functions Ø and wavelet functions Ψ by (4)

$$C_{j+1}[k,l] = \sum_{m,n} h[m-2k]h[n-2]C_j[\mathrm{m},\mathrm{n}] \tag{4}$$

The detail coefficient images are obtained from three wavelets given by Eqs. (5), (6) and (7).

$$\text{Vertical wavelet: } \Psi^1(t1,t2) = \varnothing(t1)\,\Psi(t2) \tag{5}$$

$$\text{Horizontal wavelet: } \Psi^2(t1,t2) = \varnothing(t2)\,\Psi(t1) \tag{6}$$

$$\text{Diagonal wavelet: } \Psi^3(t1,t2) = \Psi(t1)\Psi(t1) \tag{7}$$

2D DWT is used to reduce FAR and FRR. Hence improves system efficiency. Three level decomposition of normalized image of 512×64 gives the feature vector creation of size is 1×8. Energy is computed for approximation coefficients and three detail coefficients (Horizontal, vertical and diagonal) by Eq. (8)

$$\text{Energy} = \sum_{m=0}^{M-1}\sum_{n=0}^{N-1}|X(m,n)|, \tag{8}$$

where $X(m,n)$ is a discrete function whose energy is to be computed.

Table 3 FAR and FRR results and comparison

Algorithms	% FAR	% FRR	% Average accuracy
Daugman [6]	0.010	0.090	99.90
Ma et al. [7]	0.020	1.980	98
Sanchez-Reillo et al. [9]	0.030	2.080	97.89
Tisse et al. [8]	1.840	8.790	89.37
Proposed CVPR_IRIS Database (DB1)	3.021	7.800	91.44

4.4 Matching

Fuzzy hamming distance is metric given by Eq. (9) is used to compute similarity or match the value. For perfect match zero distance is shown and for any increase in distance mismatch is shown.

$$\text{Fuzzy Hamming Distance} = \frac{1}{N}\sum_{i=1}^{1}|Xi - Yi| \tag{9}$$

5 Conclusions

In this paper, imaging framework and acquisition set up is prepared to acquire good quality images, hence we have announced the availability of the CVPR_IRIS database. Also we found better accuracy for proposed database as compared with Tisse [8] as shown in Table 3.

Acknowledgements The author acknowledges the technical and nontechnical support given by CVPR Lab authorized faculty and TEQIP.

References

1. CASIA Iris Image Database. http://www.cbsr.ia.ac.cn/IrisDatabase.html
2. Proença H, Filipe S, Santos R, Oliveira J, Alexandre LA (2010) IEEE Trans Pattern Anal Mach Intell 32(8):1529–1535
3. Daugman J (1993) High confidence visual recognition of persons by a test of statistical independence. IEEE Trans Pattern Anal Mach Intell 15(11):1048–1161
4. Boles W, Boashash B (1993) A human identification technique using images of the iris and wavelet transform. IEEE Trans Signal Process 46(4):1185–1188
5. Wildes RP (1997) Iris recognition: an emerging biometric technology. Proc IEEE 85(9):1348–1363

6. Daugman J (2001) Statistical richness of visual information: update on recognizing persons by iris patterns. Int J Comput Vis 45(1):25–38
7. Ma L, Wrag Y, Tam T (2002) Iris recognition using circular symmetric filters. In: 16th international conference on pattern recognition, vol 2, pp 414–417
8. Tisse C-L, Torres L, Robert M (2002) Person identification based on iris patterns. In: Proceedings of the 15th international conference on vision interface, pp 215–229
9. Sanchez-Reillo R, Sanchez-Avila C, De Martin-Roche D (2002) Iris recognition for biometric identification using dyadic wavelet transform zero crossing. In: Proceeding of IEEE 35th Carnahan international conference on security technology, pp 272–277
10. Daugman J (2004) How iris recognition works. IEEE Trans Circ Syst Video Technol 14:21–30
11. Sanchez-Avila C, Sanchez-Reillo R (2002) Iris-based biometric recognition using dyadic wavelet transform. IEEE Aerosp Electron Syst Mag 17:3–6
12. Ma L, Tan T, Wang Y, Zhang D (2003) Personal recognition based on iris texture analysis. IEEE Trans Pattern Anal Mach Intell 25(12):1519–1533
13. Poursaberi A, Araabi BN (2007) Iris recognition for partially occluded images, methodology and sensitivity analysis. EURASIP J Adv Signal Process (Springer open J) 2007, Article ID 36751:1–12
14. Sun Z, Wang Y, Tan T, Cui J (2005) Improving iris recognition accuracy via cascaded classifier. IEEE Trans Syst Man Cybern Part C Appl Rev 35(3):435–440
15. Thornton J, Savvides M (2007) Kumar V (2007) A Bayesian approach to deformed pattern matching of iris images. IEEE Trans Pattern Anal Mach Intell 29(4):596–606

Fusion-Based Segmentation Technique for Improving the Diagnosis of MRI Brain Tumor in CAD Applications

Bharathi Deepa, Manimegalai Govindan Sumithra, Venkatesan Chandran and Varadan Gnanaprakash

Abstract Diagnosing the brain tumor from Magnetic Resonance Imaging (MRI) in Computer-Aided Diagnosis (CAD) applications is one of the challenging task in medical image processing. Traditionally many segmentation methods are used to address this issue. This paper introduces a segmentation method along with image fusion. Here a Discrete Wavelet Transform (DWT) method is chosen, for image fusion followed by segmentation using Support Vector Machine (SVM) for detecting the abnormality region. The types of MRI images considered here include T1-weighted (T1-w), T2-weighted (T2-w) and FLAIR images. The various fusion combinations are T1-w and T2-w, T1-w and FLAIR, T2-w and FLAIR. Experimental results suggest that on an average, fusion-based segmented result is superior to non-fusion-based segmented result.

1 Introduction

Magnetic Resonance Imaging (MRI) is widely used for brain tumor detection. Computer-Aided Diagnosis (CAD) succors the neuro-radiologists by giving a second opinion to improve the exactitude in diagnosing MRI brain pathology. In [1, 2], different segmentation methods with preprocessing and feature extraction have been discussed and it is also useful for the radiologist to make a decision while detecting the abnormality. The classification of abnormal regions in the given input MRI brain

B. Deepa (✉)
Department of ECE, Jayaram College of Engineering and Technology, Trichy, India
e-mail: cool.deeps.143@gmail.com

M. G. Sumithra · V. Chandran · V. Gnanaprakash
Department of ECE, Bannari Amman Institute of Technology, Sathyamangalam, Erode, India
e-mail: mgsumithra@rediffmail.com

V. Chandran
e-mail: chandran@bitsathy.ac.in

V. Gnanaprakash
e-mail: gnanaprakash@bitsathy.ac.in

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_31

image with the help of GLCM technique was explained in [3]. In [4] segmentation is carried out using spatial FCM technique for the detection of pathology. But here, fusion is used to prevent the loss of pixels without any preprocessing method in order to segment the abnormality region in the MRI brain images.

2 Related Work

Sachdeva et al. [5] segmented the tumor boundaries by identifying the weak and false edges on the brain image. Here, the hybrid methodologies such as Genetic Algorithm with Support Vector Machine (GA-SVM), and Genetic Algorithm with Artificial Neural Network (GA-ANN) were implemented for classifying the tumor region. But, this work failed to reduce the misclassification rate of tumor classification. Ng et al. [6] segmented the brain tumor by using Fuzzy C Means algorithm. Here, the membership function is defined for making different clusters. The drawback of this method is that the cluster boundaries are not well defined. Thresholding technique is most commonly used in medical image segmentation. This methodology attempts to find the threshold value for the classification of pixels into different categories like tumor and non-tumor regions [7]. Barman et al. [8] represented level set segmentation using Partial Differential Equation (PDE). K-Means is one of the simplest unsupervised algorithms that has been used for image segmentation. The given input data is clustered into different groups and the centroid (calculated by means of Euclidean distance) of one cluster is placed far away from the other cluster as possible. It fails to segment the noise corrupted images [9].

3 Proposed Work

The proposed novel technique is described in Fig. 1. It includes two steps for diagnosis of MRI brain tumor. The first step is image fusion by Discrete Wavelet Transform (DWT) and the second step is image segmentation by Support Vector Machine (SVM). The methodology used in each step has been discussed in the following sections.

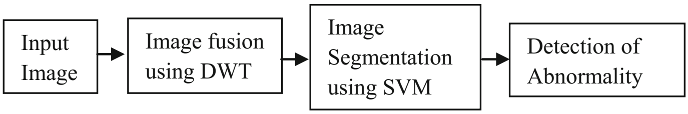


Fig. 1 Proposed methodology

In discrete wavelet transform, the given input images are decomposed into approximate (low frequency) and detail (high frequency) components [10]. Again the detail components are decomposed as horizontal, vertical, and diagonal components. The decomposition is done until the image contains the single pixel. Haar wavelet is used to decompose the image for obtaining the wavelet coefficients and maximum frequency selection rule is used for selecting the maximum absolute values for fusing the images [11]. Finally, inverse wavelet transform is used to get the fused image. Support Vector Machine (SVM) is a supervised segmentation technique used for medical images. It divides the given input fused image into resultant regions either in a rectilinear or non-rectilinear manner which is defined as a hyperplane. It is described by the training points (Supporting vector) and is given by

$$f(a) = za + q = 0 \tag{1}$$

where z is the weight vector and q is bias or threshold. If the training data is rectilinearly separable, then it is represented by $[a_i, b_i]$, $i = 1, 2, \ldots N$, where N is the number of training samples and bi belongs to $(-1, 1)$. If $b_i = 1$, it represents class 1 and if $b_i = -1$, it represents class 2 [12]. These two can be combined by the following equation:

$$b_i(z \, . \, a_i + q) - 1 \geq 0 \tag{2}$$

The kernel function $f\ (a_i, b_i)$ is demarcated by Exponential Basis Function kernel (EBF kernel), which defines the solution to map the low dimensional space vector sets to a high dimensional vector sets in case of non-rectilinear patterns [13]. The EBF kernel is given by

$$f(a_i, b_i) = \exp\left(-\frac{\|a_i - b_i\|}{2\sigma^2}\right) \tag{3}$$

4 Result Analysis

Brain images used in this paper are taken from BRATS image database. The different types of MRI brain images like T1-weighted (T1-w), T2-weighed (T2-w) and FLAIR of same patients were analyzed for the accurate detection of abnormality. After making a literature survey, here few segmentation methods such as K-Means (KM), Thresholding (ThH), Level Set (LS), Fuzzy C Means (FCM), and Support Vector Machine (SVM) are preferred, because these methods are used mostly for medical images rather than other kinds of image. The enactment of these segmentation methods considered here is evaluated with and without fusion technique.

4.1 Sensitivity, Specificity and Accuracy

Sensitivity is defined as positive probability function also termed as true positive fraction. Specificity is known as negative probability function also termed as true negative fraction

$$\text{Sensitivity} = \frac{\text{Tpo}}{\text{Tpo} + \text{Fne}} * 100\% \tag{4}$$

$$\text{Specificity} = \frac{\text{Tne}}{\text{Tne} + \text{Fpo}} * 100\% \tag{5}$$

where Tpo is the True positive, Fne is the False negative, Tne is the True negative, and Fpo is false positive. Accuracy is calculated by

$$\text{Accuracy} = \frac{\text{Tpo} + \text{Tne}}{\text{Tpo} + \text{Fne} + \text{Tne} + \text{Fpo}} * 100\% \tag{6}$$

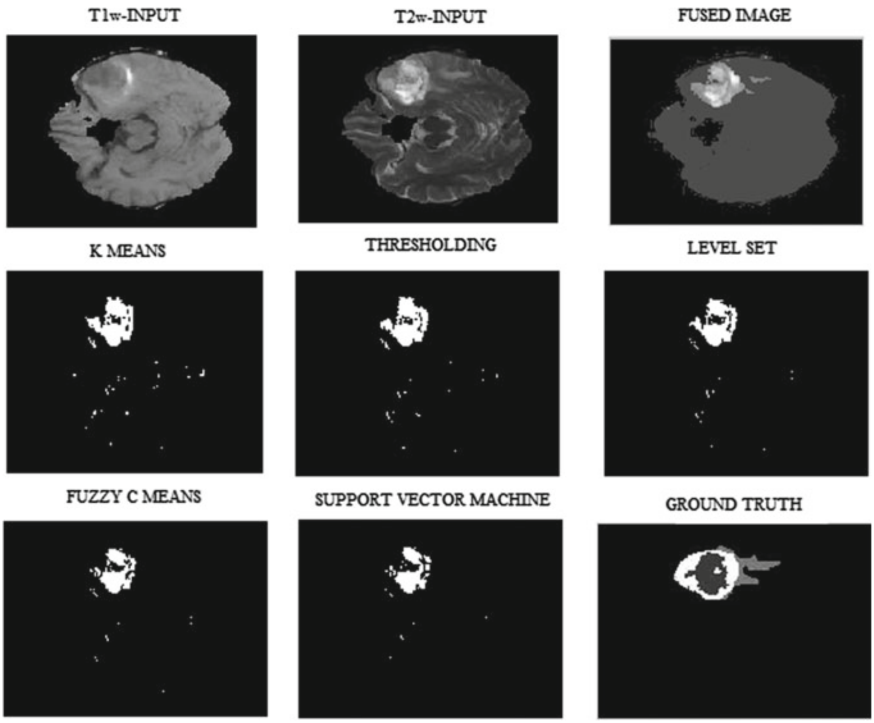


Fig. 2 Evaluation of different segmentation techniques for MRI T1-w and T2-w brain tumor image

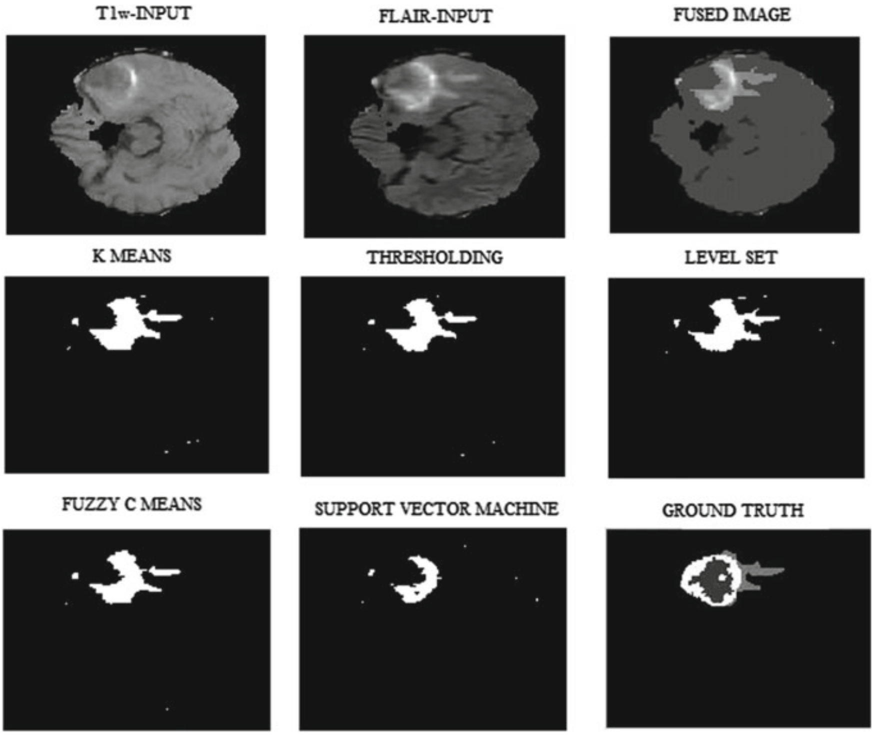


Fig. 3 Evaluation of different segmentation techniques for MRI T1-w and FLAIR brain tumor image

4.2 *Positive Predictive Value (PPV) and Negative Predictive Value (NPV)*

PPV and NPV for segmentation algorithm are calculated by

$$\text{PPV} = \frac{\text{Tpo}}{\text{Tpo} + \text{Fpo}} * 100\%; \quad \text{NPV} = \frac{\text{Tne}}{\text{Tne} + \text{Fne}} * 100\% \tag{7}$$

T1-w and T2-w image fusion with segmented result and its ground truth image (obtained from BRATS database) is shown in Fig. 2. The segmented result for this fusion image is better while using SVM rather than other methods. Figure 3 indicates the evaluation of various segmentation methods for the fusion of T1-w and FLAIR type MRI images. FCM and SVM results are good compared to others. T2-w and FLAIR fusion followed by segmentation is shown in Fig. 4. The result of SVM is better compared to all other abovementioned results considered here.

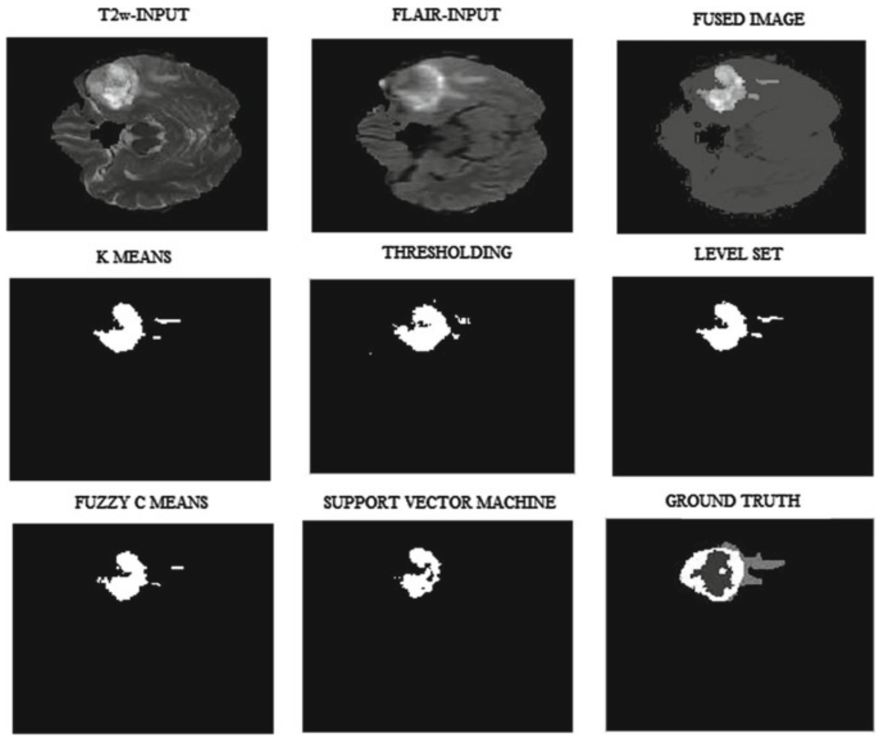


Fig. 4 Evaluation of different segmentation techniques for MRI T2-w and FLAIR brain tumor image

Figure 5 represents the performance metric for T1-w and T2-w fusion-based segmentation techniques. The obtained graphical values show that DWT fusion with SVM segmentation is giving higher result compared to others. Moreover, all the performance metric value is high in fusion-based segmentation rather than non-fusion-based segmentation. The evaluation measure for T1-w and FLAIR fused images with segmentation is shown in Fig. 6. The performance value of DWT fusion with FCM segmentation and SVM goes hand in hand. Figure 7 implicates the performance measure for fusion of T2-w and FLAIR type images followed by segmentation. It is seen that high accuracy value is for DWT fusion with SVM segmentation compared to all other techniques as discussed here. DWT with K-Means algorithm is giving inferior result.

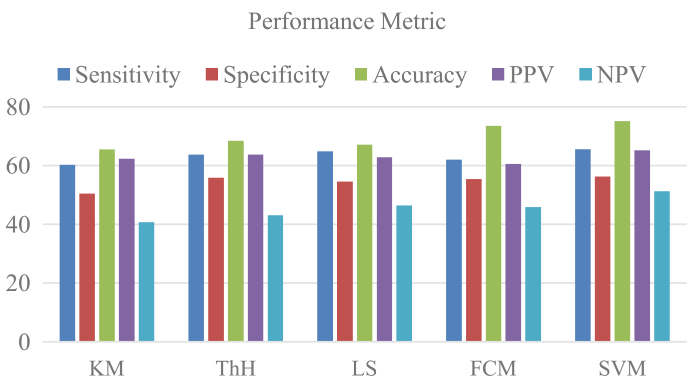


Fig. 5 Comparison of various fused segmentation techniques for MRI T1-w and T2-w brain tumor image

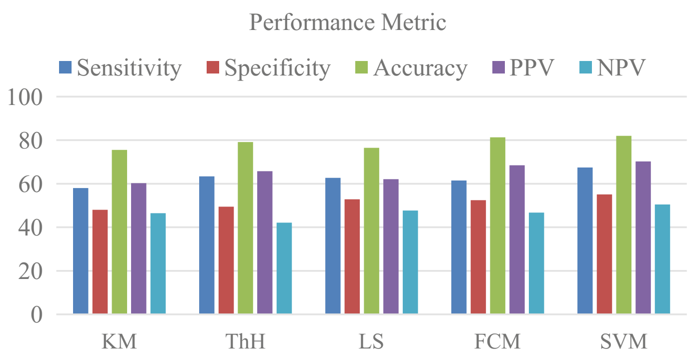


Fig. 6 Comparison of various fused segmentation techniques for MRI T1-w and FLAIR brain tumor image

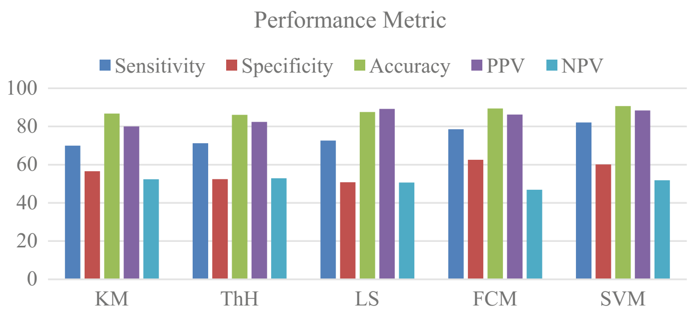


Fig. 7 Comparison of various fused segmentation techniques for MRI T2-w and FLAIR brain tumor image

Table 1 Comparitive measure of various segmentation techniques for MRI T1-w, T2-w and FLAIR brain tumor image

Input	Segmentation methods	Sensitivity	Specificity	Accuracy	PPV	NPV
T1w	KM	44.1	43.1	62.1	45.2	44.4
	ThH	44.7	42.8	63.8	44.5	45.7
	LS	43.5	44.7	63.4	46.8	46.6
	FCM	45.2	43.2	64.2	47.1	45.1
	SVM	45.9	45.7	65.5	47.8	48.2
T2w	KM	43.4	42.6	63.2	44.7	45.7
	ThH	42.5	43.9	64.8	46.2	46.4
	LS	43.8	44.7	64.6	46.9	47.5
	FCM	44.5	45.2	65.3	47.2	43.4
	SVM	45.4	46.2	66.8	48.8	46.9
FLAIR	KM	45.8	45.4	65.9	46.1	43.7
	ThH	46.1	44.4	64.6	47.5	45.4
	LS	48.7	46.5	65.7	48.2	46.5
	FCM	49.4	46.9	66.8	46.7	47.8
	SVM	50.7	48.1	67.2	49.5	45.4

Table 1 represents the performance evaluation measure like sensitivity, specificity, accuracy, Positive Predictive Value (PPV), Negative Predictive Value (NPV) for T1-w, T2-w, and FLAIR-type MRI brain tumor images by undergoing various segmentation techniques for diagnosing tumor. From the obtained tabulated values, it is seen that SVM method is giving high accuracy, sensitivity, and PPV value compared to others for all types of images.

5 Conclusion

In this paper, DWT fusion-based segmentation techniques for diagnosing brain tumor is presented. Various segmentation algorithms along with and without fusion have been examined for MRI brain images in gray scale by objective measures. From the attained results, it is seen that on an average, DWT fusion followed by segmentation methods is giving better results than non-fusion-based segmentation techniques by 20.23%. In addition, the accurate diagnosis of tumor region is achieved while using DWT fusion with SVM segmentation rather than non-fusion-based segmentation. The future work involves the introduction of novelty in wavelet transform for fusion and proposing a new hybrid segmentation technique based on intensity variation for the abnormality detection that can be finally used for classification of MRI brain tumor in CAD applications.

Acknowledgements The website link for BRATS image database is https://www.smir.ch/BRATS/Start2013. This data set was supported for my doctoral degree purpose only. We have no conflict of interest with regard to the work presented. Ethical approval to conduct this study was obtained for my research work. Informed consent was obtained from all individual participants in the study.

References

1. Abdullah N, Chuen L, Ngah U, Ahmad K (2011) Improvement of MRI brain classification using principal component analysis. In: 2011 IEEE international conference on control system, computing and engineering (ICCSCE). IEEE, pp 557–561
2. Najafi S, Amirani M, Sedghi Z (2011) A new approach to MRI brain images classification. In: 2011 19th Iranian conference on electrical engineering (ICEE). IEEE, pp 1–5
3. Singh D, Kaur K. Classification of abnormalities in brain MRI images using GLCM, PCA and SVM. Int J Eng I:243–248 (Online). http://www.ijeat.org/attachments/File/v1i6/F0676081612.pdf2012
4. Reddy AR, Prasad E, Reddy DL (2012) Abnormality detection of brain MRI images using a new spatial FCM algorithm. Int J Eng Sci Adv Technol 2(1):1–7
5. Sachdeva J et al (2016) A package-SFERCB-"Segmentation, feature extraction, reduction and classification analysis by both SVM and ANN for brain tumors". Appl Soft Comput 47:151–167
6. Ng CR, Than JCM, Noor NM, Rijal OM (2015) Double segmentation method for brain region using FCM and graph cut for CT scan imges. In: IEEE international conference on signal and image processing applications, 978-1-4799-8996-6/15
7. Sezgin M, Sankur B (2004) Survey over image thresholding techniques and quantitative performance evaluation. J Electron Imaging 13(1):146–165
8. Barman PC, Miah MS, Singh BC, Khatun MT (2011) MRI image segmentation using level set method and implement an medical diagnosis system. Comput Sci Eng Int J 1(5)
9. Liu J, Guo L (2015) A new brain MRI image segmentation strategy based on wavelet transform and K-means clustering. IEEE 978-1-4799-8920-1-15
10. Lan T, Xiao Z, Li Y, Ding Y, Qin Z (2014) Multimodal medical image fusion using wavelet transform and human vision system. ICALIP,978-1-4799-3903-9/4. IEEE
11. Indira KP, Hemamalini R (2015) Impact of co-efficient selection rules on the performance of DWT based fusion on medical images. In: International conference on robotics, automation, control and embedded systems. ISBN 978-81-925974-3-0
12. Vijayakumar B, Chaturvedi A (2012) Automatic brain tumors segmentation of MR images using fluid vector flow and support vector machine. Res J Inf Technol 4:108–114
13. Hota HS, Shukla SP, Gulhare K (2013) Review of intelligent techniques applied for classification and preprocessing of medical image data. IJCSI Int J Comput Sci Issues 1:267–272 (Online). http://www.ijcsi.org/papers/IJCSI-10-1-3-267-272.pdf

Identification of Cyst Present in Ultrasound PCOS Using Discrete Wavelet Transform

R. Vinodhini and R. Suganya

Abstract The Polycystic Ovary Syndrome (PCOS) is an endocrine abnormality; it affects females during their reproductive cycle. It is hormone imbalance of female and it skips the menstrual cycle and makes it harder to get pregnant. The side effects of PCOS are causing blood pressure, heart disease, diabetes, obesity, etc. Thus, there is an imbalance in hormone that creates many cysts in ovary and it is called as polycystic ovary syndrome. It can be diagnosed by using ultrasound scan to identify the count, size, and severity of cyst. Preprocessing the medical image is the basic and initial step for medical image processing to remove the speckle noise, present in the ultrasound image and also helpful to create medical image applications. Compared with all other modalities of scan images, ultrasound scan is less cost-effective, but it contains more speckles due to image acquisition. Speckle is a granular noise that inherently exists in and it degrades the quality of the radar, ultrasound, and CT scan images. It also causes the difficulties for image interpretation. The aim of this research paper is to apply the modified Daubechies—discrete wavelet filters for ultrasonic scan image of PCOS for removing speckle noise for better diagnose cyst. This paper concludes the effectiveness of the proposed filter to identify cyst present in ultrasound PCOS. The following metrics—SNR, PSNR, and SSIM are used to measure the effectiveness of various categories of discrete wavelet transform.

1 Introduction

Nowadays, most of the women at the age of 18–35 are affected from hormonal imbalance called PCOS. Polycystic Ovary Syndrome is a complex problem that affects 6 out of 15 women at their reproductive age. An ovary consists of follicles

R. Vinodhini (✉) · R. Suganya
Department of Information Technology, Thiagarajar College of Engineering, Madurai, Tamil Nadu, India
e-mail: vinodhinicse95@gmail.com

R. Suganya
e-mail: rsuganya@tce.edu

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_32

filled with fluid structure. The gland that is situated at the base of the brain segregates Follicle Stimulating Hormone (FSH) that can secrete, and it converts the immature follicle into mature follicle and that follicle generates estrogens, and releases matured follicle and egg at the time of menstrual cycle, on the other hand, some of that follicles are changed into chocolate cyst and due to imbalanced diet, stress and depression are formed at ovary are called polycystic ovary disorder. It causes because our blood contains a high level of insulin produced by pancreases and production of male hormone (androgen) leads to infertility. It causes other abnormalities in a woman's body, such as hair fall from the scalp, growing unwanted hair in face, chest and stomach, lack of sleep, depression, obesity and also suffers from other disease such as diabetes, heart diseases and high blood pressure. It can be diagnosed by calculating Body Mass Index (BMI) of a patient with periodical review along with blood pressure, hormone level test, and ultrasonic scan to find out the size, count and severity of cyst. The purpose of this research paper is to identify these cysts at the earlier stage from the ultrasonic images by using various filtering techniques, compare them to find out which should be suitable for PCOS disorder. Filtering is the core and initial step used in image processing to enhance and remove noise from the image. It needs a higher level of decomposition filter to remove noise. The presence of speckle in a medical image would hide the necessary pathological information leading to improper result in a diagnosis.

2 Related Works

The related works narrated the importance of decomposition filters of ultrasonic images: [1]. The author proposes that anisotropic diffusion filters are used to remove the speckles of the ultrasound scan images and radar images for edge diffusion [2]. The speckled SAR and ultrasonic images are difficult to analyze due to the presence of more speckle noise, and the author uses the discrete wavelet transform filter to remove speckles of the image. In [3], the author proposes a speckle suppression technique to ultrasonic image for preprocessing [4]. The author proposes texture classification technique to differentiate the normal liver and fatty liver by using Discrete Wavelet Transform. The author [5] proposes to analyze the quality of ultrasonic image by applying various DWT filters for denoising the image by using the soft thresholding and compare the DWT filters by PSNR, SNR ratio. This paper [6] aims to compress the ultrasonic medical image based on the Region of Interest (ROI) for analyzing the patient record by using wavelet transform and their quality is measured by the metrics SNR and SSIM [7]. The author says that the quality of their ultrasonic's image is measured by using signal-to-noise ratio and they can be in the range of 8.5, 5.5, 2.2. PCOS is the endocrine disorder that affects women at the age of 18–35, and in paper [8], the author analyzes the disease by k-means algorithm and classify the number of follicles in ovary. PCOS is diagnosed by the medical record of the patient along with the ultrasound image of the abdomen of the ovary [9]. The author proposed the novel method to automate and detect the PCOS detection [10]. The excessive segregation of the insulin in the female hormone creates the segregation of androgen (male hormone). The presence of the cyst (immature follicle) in the ovary is diagnosed easily by the ultrasound image.

But all the above paper fails to address the subtle features of speckle noise contain some important information of immature follicle which causes infertility in the future. So, we planned to identify this subtle features present in the noisy ultrasound image rather than eliminating a noise.

3 Proposed Work

Automatic interpretation of ultrasound images, however, is extremely difficult because of its low signal-to-noise ratio (SNR) [11]. Image preprocessing is a technique to enhance a medical image. The ultrasonic scanning image of a patient affected from Polycystic Ovary Syndrome and uses filtering techniques to remove and analyze speckles in the image. Speckle noises tend to obscure diagnostically important features and degrade the image quality and thus increase the difficulties in diagnosis. This work is based on the major issue in ultrasound modality. This work provides knowledge about adaptive and anisotropic diffusion techniques for speckle noise removal from various types of ultrasonic images. Though much work has been done on the speckle reduction in ultrasound images in the literature, in most of the work, it is assumed that the Diffusion-based spatial filter is to suppress the noise and preserve fine details of edges in ultrasound images. But it looks to diagnose small residual features like cyst and lesion which will lead to major problems like cancer and finally lead to death. So much care should be taken to identify small pathology bearing information which is hidden by means of speckles during preprocessing. In order to focus on subtle features like immature follicles, chocolate cyst and improve edge information for ultrasound PCOs images. In this research work, we have proposed the db filter for preprocessing technique that eliminates only speckle noise and leaves subtle features which help to identify the presence of cyst or immature follicles in its earlier stages (Fig. 1).

In this proposed work, pre-process the image at the size of 222 KB and JPEG format image. From the above survey, DWT (haar, bior, db) and also with continuous wavelet filters are applied to ultrasonic image, compare the results of that image and identify which filter is suitable for filtering ultrasonic image. Comparison of filters is among metrics such as SNR, PSNR, and SSIM ratio.

4 Methodology Used

(a) Discrete Wavelet Transforms

To remove speckles from medical image, there is need of a number of decompositional layers from the survey about the filters used in ultrasonic image the above two

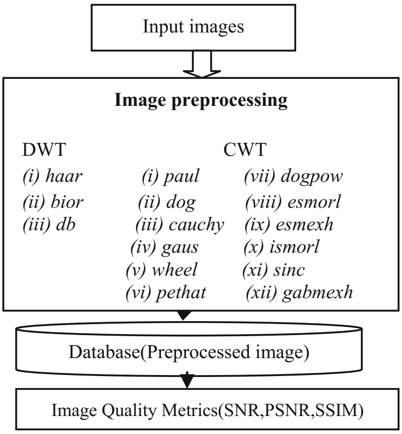


Fig. 1 Block diagram of the proposed work

filters are mostly used: Discrete Wavelet Transform (DWT) and Continuous Wavelet Transform/anisotropic filter (CWT).

DWT:

The Discrete Wavelet Transform (DWT) has turned into an intense procedure in biomedical signal processing. The most evident distinction is that the DWT utilize scale and position esteems in light of forces of two. The values of s and τ are: $s = 2^j$, $\tau = k * 2^j$ and $(j, k) \in Z^2$ as shown in Eq. (1).

$$\psi_s, \tau(t) = \frac{1}{\sqrt{2^j}} \psi\left(x = \frac{t - k * 2j}{2^j}\right) \quad (1)$$

The essential thought behind disintegration and remaking is low-pass and high-pass separating with the utilization of down examining and up inspecting individually. The consequence of wavelet decay is progressively composed disintegrations. In this proposed work, we connected the subchannels of the DWT with the [1] delicate thresholding to denoise the picture.

(b) **CWT/Anisotropic filter**

The continuous wavelet transform (CWT) selects the interior eventual outcome of a pennant, $f(t)$, with deciphered and created varieties of an analyzing wavelet, $\psi(t)$. The logical state of CWT is:

$$C(a, b; f(t), \psi(t)) = \int f(t) \frac{1}{a} \psi * (t - b/a) \mathrm{d}t \quad (2)$$

where $f(\omega)$ and $\psi(\omega)$ are the Fourier transforms of the signal and the wavelet.

We also used an anisotropic filter because it is also one of the decomposition filters and all the sub-filters of the CWT are used such as paul, dog, cauchy, gaus, wheel, pethat, dogpow, esmorl, esmexh, gaus2, gaus3, isodog, isomorl, sinc, and gabmexh.

(c) **Daubechies**

Speckles are the noise which are present at the medical image that degrades the quality of the image like ultrasonic and radar images. There are many filters to remove the noise of the image like low-pass filter, high-pass filter, adaptive filter, and Gabor filter but the ultrasonic image that contains the speckles which has the more noise and it needs some decomposition layer to remove the filters. Wavelet filter is one type of the decomposition filter. Recently, wavelet transform is used in the analysis of the image, it separates the speckles and signals into the different orientation and scales. Daubechies is the one of the sub-filter of the discrete wavelet transform. From the computed metric ratio in Table 1, db2 gives better solution compared with the other filters. It is an orthogonal wavelet filter and they are defined by the DWT. It is characterized by the maximal number of vanishing moments.

$$2^A - 1 \text{ (algebraic equation of db filter)} \tag{3}$$

where A is the number of highest vanishing moments.

It can be used by the two schemes they are using length (DA) and refer with vanishing moment (db2).

The db filter has the overlapping windows, so the result reflects the pixel intensity and it is smoother than haar filter, but it is more complex than haar. It has four wavelets and scaling coefficients. The sum of the functions of coefficient is calculated

Table 1 Comparison of various filters by metrics of ultrasound image

Filter name	Subname of filter	PSNR	SNR	SSIM
DWT	bior1.5	21.66	11.41	0.88
	db2	21.73	11.66	0.87
	haar	21.60	11.51	0.88
CWT	paul	21.59	4.71	0.82
	dog	20.59	4.79	0.91
	cauchy	21.67	1.74	0.98
	gaus	20.29	3.16	0.95
	wheel	20.70	5.79	0.93
	pethat	20.20	1.0	0.90
	dogpow	20.47	3.99	0.91
	esmorl	20.27	1.23	0.93
	esmexh	20.41	2.17	0.91
	isodog	21.52	6.15	0.98
	isomorl	20.18	11.17	0.88
	sinc	21.46	7.08	0.99
	gabmexh	20.42	1.79	0.97

by the average over the adjacent pixels increase its capabilities of image processing techniques.

4.1 Image Quality Metrics

Peak signal-to-noise ratio (PSNR):

This measure is used to compare the square error between the original and constructed image. PSNR using the term Mean Square Error in the denominator, so the PSNR rate is high means image has low error.

$$\text{MSE} = \frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}[I(i, j) - K(i, j)]^2 \tag{4}$$

The PSNR (in DB is defined as)

$$\text{PSNR} = 10\,.\,\log_{10}\left(\frac{\text{MAX}_I^2}{\text{MSE}}\right) \tag{5}$$

where MAX_1, Maximum possible pixel value of the image. MSE, Mean Square Error.

High PSNR means that the image has good quality and less error. Range of PSNR for ultrasound image is at 20.45, 20.72, 21.90.

Signal-to-noise ratio (SNR):

It is used to measure the sensitivity of the image. SNR is the ratio of average signal value μ_{sig} to the standard deviation of the signal $\sigma_{\text{sig.}}$

$$\text{SNR} = \frac{\mu_{\text{sig}}}{\sigma_{\text{sig}}} \tag{6}$$

Structural similarity index measurement (SSIM):

It is used for prediction of image quality is based on an initial uncompressed or distortion-free image as reference. The SSIM of the image is calculated on the various window sizes of the image. They measure between the two windows are x and y

$$\text{SSIM}(x, y) = \frac{\left(2\mu_x\mu_y + c1\right)\left(2\sigma_{xy} + c2\right)}{\left(\mu_x^2 + \mu_y^2 + c1\right)\left(\sigma_x^2 + \sigma_y^2 + c2\right)} \tag{7}$$

where

μ_x, μ_y the average of x and y, σ_x^2, σ_y^2 the variance of x and y
σ_{xy} the covariance of x and y and $c1 = (k_1L)^2$, $c2 = (K_2L)^2$

5 Result and Discussion

The experimental data of this paper are obtained from the radiology department of Meenakshi Mission Hospital in Madurai, India. We have collected nearly 20 women scan reports and conducted a small survey. Preprocessing is the root process for medical image processing and other applications by imaging. In this paper adapted to filter techniques to remove noise from an ultrasound scan image for PCOS. These can be taken of the abdomen of the patients whom are affected from PCOS periodically.

Figure 2 represents the ultrasonic scan image of the PCOS patient and using this, one can diagnose the presence of cyst (immature follicles) in the ovary and find out the severity of the disease (Fig. 3).

The Daubechies filter that it is a type of the Discrete wavelet filter is used to remove speckles with that fix the decomposition layer of filter at seven because this ultrasound image needs more number of the decomposition layer to increase the PSNR ratio, if the PSNR ratio is high the image has good quality. Comparison of the image with different decomposition layers is represented below the graph.

Fig. 2 PCOS ultrasonic image

Fig. 3 db filter with decomposition layer

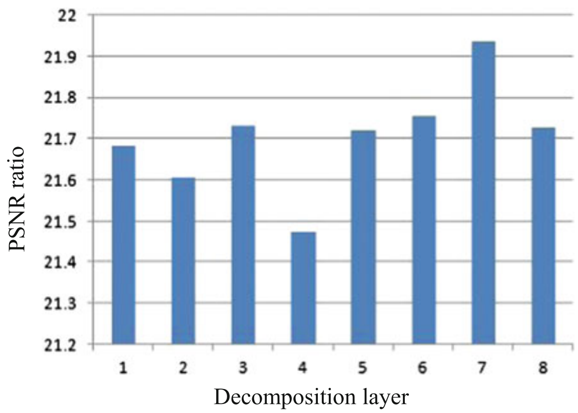
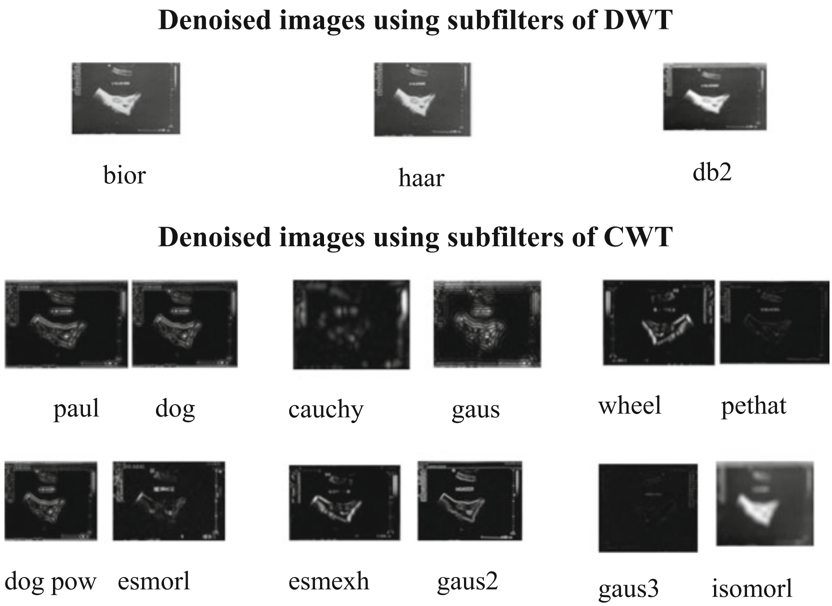


Fig. 4 PSNR value for different decomposition layer

Figure 4 represents the ultrasound image with the different level of decomposition layer, and we have considered the decomposition layer as seven for effective preprocessing, it has high value of the PSNR ratio (21.9). Approximation level, horizontal, diagonal and vertical details of the image are also determined.



Above figures represent the denoising image of the various subfilters of wavelets. From this, we have chosen a suitable filter. It can be measured by using image quality metrics—PSNR, SNR, and SSIM. The comparison of PSNR, SNR, and SSIM for the above-filtered image is as follows.

Various filters of DWT and CWT are applied for PCOS dataset and measured by the three metrics (PSNR, SNR, SSIM): high PSNR range shows that the image has higher quality in Table 1 represents the comparison of all filters, db2 has high in ratio (21.7316). The range of SNR for ultrasound image is (2-12) and db2 also satisfies the SNR ratio as 11.6566. The range of SSIM ratios for ultrasound image is between −1 and 1, and SSIM value for db2 is 0.8673. From this, we have concluded that db2 satisfies all the metric ratio—PSNR, SNR, and SSIM. So, db2 is best for removing speckles and retains subtle features of PCOS ultrasonic image information.

6 Conclusion

The most important technique to diagnose PCOS for women at the earliest stage is ultrasound scan modality. In medical imaging, preprocessing is the initial step to remove speckle noise, and various subfilters of DWT and CWT with different decompositional layers are experimented to remove the speckles for retaining the subtle feature. Noise ratio of preprocessed image is also calculated by image quality metrics—PSNR, SNR, and SSIM. By diagnosing the preprocessed image, db2 is examined as a suitable filter for ultrasound image speckle reduction.

References

1. Yu Y, Acton ST (2002) Speckle reducing anisotropic diffusion. IEEE Trans Sig Process Soc 11(11):1260–1270
2. Choi H, Jeong J (2018) Despeckling images using a preprocessing filter and discrete wavelet transform-based noise reduction techniques. IEEE Sens Counc 99:1
3. Hao X, Gao S, Gao X (1999) A novel multiscale nonlinear thresholding method for ultrasonic speckle suppressing. IEEE Trans Med Imaging 18(9):787–794
4. Huang Y, Wang L, Li C (2008) Texture analysis of ultrasonic liver image based on wavelet transform and probabilistic neural network. In: IEEE conference
5. Thakur A, Anand RS (2005) Image quality based comparative evaluation of wavelet filters in ultrasound speckle. Digit Sig Process 15(5):455–465
6. Gupta S, Bhatia R (2007) Comparative analysis of image compression techniques: a case study on medical images. In: IEEE conference on advances in recent technologies in communication and computing
7. Omar EA, Varghese T (2011) Signal to noise ratio comparisons for ultrasound attenuation slope estimation algorithms. Nucleic Acid Res
8. Dewailly D et al (2017) Diagnosis of polycystic ovary syndrome (PCOS): revisiting the threshold values of follicle count on ultrasound and of the serum AMH level for the definition of polycystic ovaries. Oxford 26(11):3123–3129
9. Lawrence MJ, Eramian MG, Pierson RA, Neufeld E (2007) Computer assisted detection of polycystic ovary morphology in ultrasound images. In: IEEE conference computer and robot vision, 2007

10. Padmapriya B, Kesavamurthy T (2015) Diagnostic tool for PCOS classification. In: Springer conference, vol 52
11. Ramamoorthy S, Subramanian RS, Gandhi D (2014) An efficient method for speckle reduction in ultrasound liver image for e-health applications. In: International conference on distributed computing and internet technology, vol 8337, pp 311–321

Design and Development of Image Retrieval in Documents Using Journal Logo Matching

S. Balan and P. Ponmuthuramalingam

Abstract This research focuses on the detailed study of document image retrieval form World Wide Web using the existing techniques, text and image information which are extracted in the database. Retrieval based on the user's given keywords of journal papers. It fetches the information from the extracted folder and analyzing with logo matching technique using efficient SURF (Speeded-Up Robust Features) method. Precision, recall, f-measure, time, and memory are calculated to compute the accuracy of the given input. Logos are clustered using image descriptor methods using shape, edge, and color techniques. This paper proposes a new technique for logo matching in journal database depends upon the documents.

1 Introduction

Document image caries lot of text, printed, or scanned images. In the current scenario, web store tons of documents in different formats. Each format varies type and size. So the storage of information is to be text or image. Here, the process is based on journal documents (Word and Pdf) information retrieval from the web is extracted and preprocessed. In past, surveys on document images are focused on text conversion analysis, character images of text, indexing the structure of document, image abstraction via automatic, GIF images, drawings, and semantic image relations [1].

A document page is captured by using pixel-level processing method. It is separated by number of characters in the document and split into two different lines. Region values are identified by using feature values of the text and graphics of image in the given document description. For each image, binary value and threshold value are identified by using existing techniques. Some portion of text in the image is dark,

S. Balan (✉) · P. Ponmuthuramalingam
Research Department of Computer Science, Government Arts College (Autonomous), Coimbatore, Tamil Nadu, India
e-mail: Balan.sethuramalingam@gmail.com

P. Ponmuthuramalingam
e-mail: ponmuthucbe@gmail.com

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_33

Fig. 1 Sample IEEE logos for feature extraction

low, printed, same image with less or higher quality, extracted from we, .doc, .pdf, .HTML. So the size and clarity of image differs. Same matching image is identified by using the threshold value.

Here, Fig. 1 shows some of the sample logos of journal database is shown below. In that, IEEE images are extracted from the documents and cluster the images based on feature extraction values. For example, the logos are taken from IEEE website, Conference Template, IEEE conference advertisement, Journal Documents and Google search. This paper is categorized as follows: (1) study on Document Image retrieval (2) Survey of text and retrieval methods (3) using SURF methods (4) of performance evaluation (5) further research directions in logo matching is stated.

2 Literature Review: Survey

This survey aims to solve the various document image retrieval techniques to achieve the accuracy of the given documents. Document Image Retrieval (DIR) is used to identify the relevant documents based on the feature of the image. Documents stored in web are normally as E-books and handwritten documents. The page layout structure and text searching in documents is said to be Digital Libraries (DL). Three components of DL are as follows: indexing, querying, and similarity [2].

For example, processing step of a book is representation (paper, descriptive metadata, pixel, structural metadata, feature and abstract repository), indexing (Database Management System, Image Database, DIR Database, and Information retrieval Engine), and Retrieval task (manual browsing, visual browsing, layout-based DIR, keyword and Full-Text). Automatic Indexing of Document Images (AIDI) describes the font indexing and document page layout structure [2].

The block diagram of DIR is query image (based on the user's query retrieves from the relevant document image), removal of noise (text image extraction such as signed documents, logo), computing feature (document image is extracted and stored in the database to represent the image attributes and values), and matching algorithm (Similarity value is compared with other images in database) [3].

Document page layout is categorized into geometric, Functional and Semantic Descriptions. Execution of the page layout is processed by Structure independent of (column, margins), logical-type dependent (address, signature), content is presentational (font, size, and style), and linguistic (meaning of block). Multiple retrieval query images are classified in the various methods: Average Query (Joint-Avg), Support Vector Machine (Joint-SVM), Average and maximum of multiple queries (AVG-MAX), and spatial re-ranking [4].

Document layout analysis based retrieval is grouped into two approaches namely top-down and bottom-up. Local content feature is based on local, global representation, and encoding. The real-time retrieval of document uses global description of image, similarity, relevance feedback, and score. The classification of document image indexing is divided into two ways: traditional indexing (Keyword spotting, OCR, New Method Signature based, Layout Structure and logo Matching) [5].

Signature matching in document is based on layout analysis, signature and handwritten text localization, binarization, noise removal, normalization, abstraction, feature value extraction, verification, and identification. Signature matching is problem of finding the similar matching from samples of signature stored in a database. It is based on value measures, detection and segmentation, structural saliency and contour grouping, and shape matching. Automatic extraction of signature retrieval based on two images namely document image signature extraction (preprocessing, finding candidate signature, extraction, and detection of feature value) and feature matching on document image retrieval on signature and document ranking [6].

Classification method used in Document Image Retrieval are KNN, ANN, DTC, MLP, CNN, and SVM. There are various different distance metrics are used as follows Canberra, Euclidean, City block, Chebychev, Cosine, Hamming and Jaccard. Indexing of documents based on signature is scanned image, segmentation, patch level classification, computing feature of signatures, touching removal and region detection [6].

Digital library consists a huge amount of documents stored in digital formats. It aims to solve the data available in the Internet is electronically categorized. The advantages of DL are physical boundary, information retrieval, preservation and conservation, resource sharing, and multiple access. DIR is based on recognition (converting document image into text) and recognition free (explicit recognition of image in relevant document) [7]. The challenges in DIR are the unavailability of OCR, degradation, poor quality, noisy images, paper quality, folding the paper, speed, cross-lingual and representation of font and encoding to find the distance matching documents in OCR uses standard distance functions (Euclidean, Manhattan, Mahalanobis, Dynamic time wrapping [8].

OCR is categorized into several parts such as character segmentation, feature extraction, and classification. Graphical document recognition mostly carries graphical information (maps and drawings) and graphical recognition (music, 2D and 3D views). OCR difficulties in graphical documents are character recognition and segmentation, separation of text, and individual text extraction. Difficulties in graphical documents area character recognition and segmentation, separation of text, and individual text extraction.

Logo similarity matching for negative shape features is identified by logo, triangular, rectangular, square, stripes, and borders. Context dependency system solves the logo matching technique consists of preprocessing the image components, brightness the values of image, improve the visual quality and adjusting the brightness values of image [9]. Some of the existing methods and dataset are used to achieve the results in the current survey as follows (Table 1).

Feature value of extraction is carried out in three ways namely color, text, and edge. Existing methods of logo spotting are based on approximate nearest algorithm, FLANN (fast library for approximate nearest neighbor), indexing, key points (query and document image) and learning. Document retrieval system query on logo matching is focused on extraction, match suing ANN, grouping in geometric filter and ranking. In document database flow is on text or non-text separation from a cluster of documents to identify the logo position [9].

Table 1 Recent survey on logo matching retrieval datasets

Authors	Method	Year	Dataset
Vaijinath et al. [9]	GLCM, ANN	2017	Organization logo
Romberg et al. [10]	Querying triangle and Edge Index	2011	FedEx logo
Dikey et al. [1]	Feature extraction	2015	Popular logo
Boonroda et al. [11]	SURF	2015	Product logo
Billa et al. [12]	Multiple descriptors	2017	Popular logo
Bharathidevi et al. [13]	SIFT descriptors	2017	Popular logo
Sawalkar et al. [14]	SIFT, RANCSAC	2014	Given images
Ramachandran et al. [7]	SURF	2015	Pair of logo

3 Proposed Evaluation of Approximate Surf Method

SURF method is used for feature detection, feature description, and feature matching. Mainly focus on integral matching and Hessian matrix. The main use is box filter techniques. It is divided into three ways namely interpoint detection, interest point description, and interest point matching.

$$\mathcal{H}(x,\sigma) = \begin{bmatrix} L_{xx}(x,\sigma) & L_{xy}(x,\sigma) \\ L_{xy}(x,\sigma) & L_{yy}(x,\sigma) \end{bmatrix}, \tag{1}$$

where L_{xx} and L_{yy} are approximate box filters

To determine good approximation values:

$$\det\left(H_{\text{approx}}^{\text{SURF}}\right) = \hat{L}_{xx}\hat{L}_{yy} - \left(0.9\hat{L}_{xy}\right)^2 \tag{2}$$

To determine both values

To calculate sum of the pixels,

$$S = W - X - Y + Z \tag{3}$$

Local Feature Detection Based on Interest Point

Algorithm is as follows:

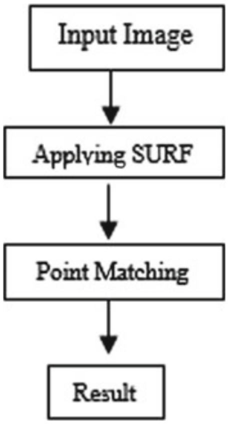
Step 1: given input image x
Step 2: point initializing
X – Filtering the feature values
For y → {0, 1, 2, and 3} do
End for
For z: = 0 to 3 do
For m: = 1 to 3 do
X → 2° m + 1
End for
End for
Return point matching

Figure 2 explains the general block diagram of SURF method is shown below:

3.1 Approximate SURF Method

Logo matching is based on text retrieval and image retrieval. Text retrieval is defined by document style and image retrieval is identified by using the approximate SURF method. Simple steps carried out for in journal logo database as follows:

Fig. 2 SURF method



Step 1: Searching the document
Step 2: extracted and stored in separate folder
Step 3: compared and cluster the feature value within text document
Step 4: image is split into a $\times$ b windows, compute the deviation
Step 5: convert into binary image and identify the threshold value
Step 6: extract the shape, boundary, edge detection regions
Step 7: image properties are stored in separate folder
Step 8: for $n = 3$: size
For $x = 1$ to n
Find logo in extracted documents
End for
: set $m = 1$ to max $= 6$
Step 9: apply the threshold value in logo matching
Step 10: matching results are displayed
Step 11: end

4 Experimental Analysis and Discussions

To measure the effectiveness of information retrieval carried out in three steps namely: gathering documents, testing the needs based on queries, and finding a set of pairs for relevant or nonrelevant documents.

4.1 Precision, Recall and F-Measure

Precision is defined by the correct accuracy of the information system.

$$P = \text{tp}/(\text{tp} + \text{fp}) \quad (4)$$

Recall is defined by the extraction of relevant information from the system.

$$R = \text{tp}/(\text{tp} + \text{fn}) \quad (5)$$

F-measure is defined by precision and recall are opposite to each other.

$$F = 2PR/P + R \quad (6)$$

Table 2 shows that the performance evaluation of logo matching and Table 3 displays the overall measure of the user's input.

Figure 3 shows the overall measure of calculating the precision, recall, and f-measure values from ACM, Springer, IEEE, and Elsevier. Table 3 shows the performance of given input query (Fig. 4).

Here, the above measure and performance are calculated by using the given users query as automatic web page logo detection. Based on this query, it fetches the document from the web and extracted to separate folder. Using the existing techniques, text retrieval is measured and for image retrieval based on logo matching using efficient SURF method. Figure 5 shows that the probability of text retrieval for the given input. From web and matches with pair of documents.

Table 2 Performance evaluation of logo matching

Logo	Precision	Recall	F-measure	Time (ms)	Memory (Bytes)
ACM	33.58	33.15	33.37	230	235,520
Springer	37.70	38.67	38.18	141	144,384
IEEE	**41.86**	**42.88**	**42.36**	**89**	**91,136**
Elsevier	39.00	39.71	39.35	101	10,324

Table 3 Performance Measure

Text	Logo	Probability
Automatic web page logo detection	ACM	5.1
	IEEE	**7.4**
	Springer	1.8
	Elsevier	1.8

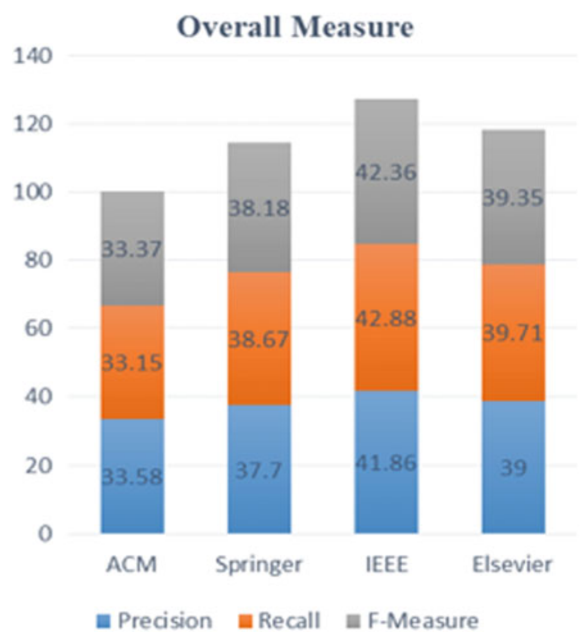


Fig. 3 Overall measure

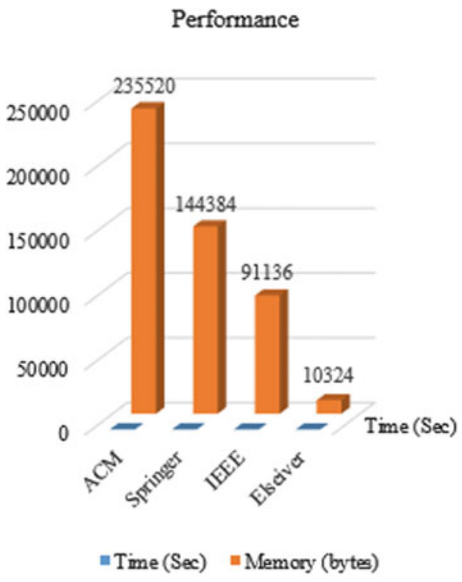
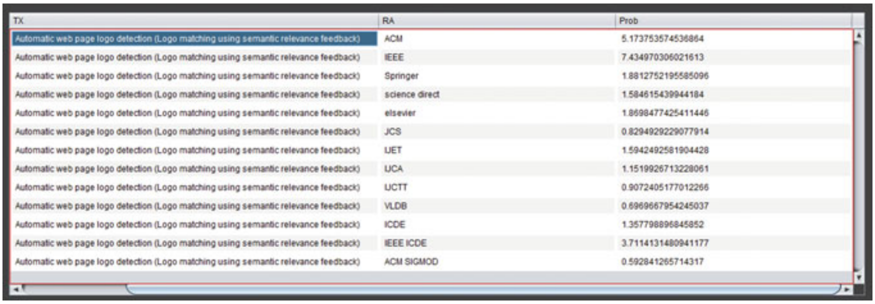


Fig. 4 Performance



TX	RA	Prob
Automatic web page logo detection (Logo matching using semantic relevance feedback)	ACM	5.173753574536864
Automatic web page logo detection (Logo matching using semantic relevance feedback)	IEEE	7.434970306021613
Automatic web page logo detection (Logo matching using semantic relevance feedback)	Springer	1.881275219558509 6
Automatic web page logo detection (Logo matching using semantic relevance feedback)	science direct	1.58461543994418 4
Automatic web page logo detection (Logo matching using semantic relevance feedback)	elsevier	1.8698477425411446
Automatic web page logo detection (Logo matching using semantic relevance feedback)	JCS	0.8294929229077914
Automatic web page logo detection (Logo matching using semantic relevance feedback)	IJET	1.5942492581904428
Automatic web page logo detection (Logo matching using semantic relevance feedback)	IJCA	1.1519926713228061
Automatic web page logo detection (Logo matching using semantic relevance feedback)	IJCTT	0.9072405177012266
Automatic web page logo detection (Logo matching using semantic relevance feedback)	VLDB	0.6969667954245037
Automatic web page logo detection (Logo matching using semantic relevance feedback)	ICDE	1.357798896845852
Automatic web page logo detection (Logo matching using semantic relevance feedback)	IEEE ICDE	3.7114131480941177
Automatic web page logo detection (Logo matching using semantic relevance feedback)	ACM SIGMOD	0.592841265714317

Fig. 5 Text retrieval for user given query

5 Conclusion and Future Scope

This paper solves the outcome of the detailed survey in document image retrieval and extraction methods. The logo matching method is based on efficient SURF techniques. Design and analysis of techniques are processed by stepwise manner. Retrieval techniques are not only focusing the text but also finding the feature extraction cluster values of logo. The probability of given query is analyzed and tested using software prototype is found successful to achieve the accuracy of documents.

Logo image database such as ACM, Springer, Elsevier and IEEE are taken to find the precision, recall, and f-measure values are identified. Compared to other logos, the extraction from IEEE database fetches the results in sort time. The retrieval time taken as 89 ms and space allotted to store the information is 91,136 bytes. In future, the retrieval techniques of logo matching focus on the whole journal database.

References

1. Dikey KPJ (2015) Review of logo matching & recognition system based on context dependency. Int J Innov Res Comput Commun Eng 3(1):281–286
2. Keyvanpour M, Tavoli R (2013) Document image retrieval: algorithms, analysis and promising directions. Int J Softw Eng Appl 7(1):93–106
3. Marinai S, Marino E, Soda G (2007) Exploring digital libraries with document image retrieval. In: Research and advanced technology for digital libraries, pp 368–379
4. Gao H et al (2015) Focused structural document image retrieval in digital mailroom applications
5. Tavoli R (2012) Classification and evaluation of document image retrieval system. WSEAS Trans Comput. E-ISSN: 2224-2872
6. Schulz T, Sablatnig R (2015) Signature matching in document image retrieval. In: 20th computer vision winter workshop
7. Ramachandran R, Jose A (2014) Logo matching and recognition system using surf. IJRCCT 3(9):981–986
8. Roy PP, Lladós J, Pal U (2011) Multi-oriented and multi-scaled text character analysis and recognition in graphical documents and their applications to document image retrieval. Universitat Autònoma de Barcelona

9. Bhosle VV et al (2017) Multi classification technique for logo based document verification. Int J Comput Eng Appl XI(XI):1–12
10. Romberg S et al (2011) Scalable logo recognition in real-world images. In: Proceedings of the 1st ACM international conference on multimedia retrieval. ACM
11. Boonrod T, Jareanpon C, Chomphuwiset P (2015) The comparison of template matching and SURF for logo classification on product. In: Proceedings of the 3rd IIAE international conference on intelligent systems and image processing, pp 256–263
12. Billa P, Balijepalli AK, Rama Koteswara Rao I (2017) An implementation of effective logo matching and detection using multiple descriptors to enhance the resolution. Int J Comput Appl 161(5)
13. Bharathidevi B et al (2017) Logo matching for document image retrieval using SIFT descriptors. Int J Eng Res Appl 7(2):55–60
14. Sawalkar MS, Patole M (2012) Logo detection and recognition from the images as well as videos. Int J Sci Res (IJSR) 3:1233–1238

Mr. S. Balan received his Master of Computer Science from Dr. G.R.D College of Science, Coimbatore in 2009 and M.Phil. in Computer Science from Bharathiar University, Coimbatore in 2012. He is pursuing his Doctor of Philosophy in Computer Science, Government Arts College (Autonomous), Coimbatore, Bharathiar University. His research interests include Data Mining, Cloud Computing, Networks and Image Processing.

Dr. P. Ponmuthuramalingam received his Master's Degree in Computer Science from Alagappa University, Karaikudi in 1988, M.Phil. in Computer Science from Bharathidasan University, Tiruchirappalli, and Ph.D. in Computer Science from Bharathiar University, Coimbatore. He is working as Controller of Examination and Associate Professor in Computer Science, Government Arts College, and Coimbatore since 1989. His research interest includes Text mining, Semantic Web, Network Security, and Parallel Algorithms. He has contributed more than 50 papers in various National/International conference/journals.

Feature Enhancement of Multispectral Images Using Vegetation, Water, and Soil Indices Image Fusion

M. HemaLatha and S. Varadarajan

Abstract Land cover characteristics of satellite images are analyzed in this research paper. Remote sensing indices are calculated for multispectral image. In the proposed method, satellite image indices, i.e., NDVI (Normalized difference vegetation index), NDWI (Normalized difference water index), and BSI (Bare soil index), are calculated for various classes such as land, vegetation, water, and in land cover categories. All these remote sensing indices are fused to get composite bands and to enhance all features in multispectral image. This technique increases visual perception of human eye for multispectral images. Fusion plays vital role in remote sensing and medical images interpretation. In case of remote sensing, we cannot get entire information in one spectral band. So multispectral bands are combined, which leads to feature enhancement. This method depends on green (G), infrared (IR), near infrared (NIR), and short wave infrared (SWIR) bands and their fusion. Finally, error matrix is generated with reference data and classified data. The main application is to calculate vegetation, bare soil, and water indices in three land covers and to get better feature enhancement. Producer's accuracy, consumer's accuracy, commission, omission, kappa coefficient, F1score, over all accuracy, and over all kappa coefficients are calculated.

1 Introduction

Advanced ecosystems and socio economic managements are monitored by remote sensing and geographical information system. Thematic maps and general purpose maps will provide some useful information. The use of remote sensing and aerial photography has made it possible to map large areas for resource management and

M. HemaLatha (✉) · S. Varadarajan
ECE Department, S.V.U.C.E, Tirupati, A.P., India
e-mail: maddihemalatha@gmail.com

S. Varadarajan
e-mail: varadasouri@gmail.com

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_34

exploitation. Quantitative spatial variation in handling huge data requires appropriate tools to analyze the spatial data using statistical methods or time series analysis [1].

Satellite classification is used to classify the ground data. There are many classification methods such as wavelet based, various cluster techniques, principle component analysis, and artificial neural networks. Satellite portraits are used for several applications including land cover and land use monitoring [2, 3] disaster management, soil moisture detection, snow and glacier monitoring, agriculture management, and forest fire monitoring [4, 5].

In this research paper, satellite portrait is divided into multi-classes using remote sensing indices. Various qualitative parameters are calculated. NDVI, NDWI, BSI, Producer's accuracy and user's accuracy plays very important role in this classification. Many texture-based classifications have been proposed [6, 7]. But there is problem in identifying water bodies. In the proposed technique, water bodies are identified. NDVI, NDWI, and BSI raster's are combined to get feature enhancement. Then accuracy assessment is done by error matrix. Producer's accuracy, kappa coefficient [8], omission, commission, F1score, over all accuracy, and over all kappa coefficient are calculated. Classification is done based on various remote sensing indices [9, 10].

2 Materials and Methods

Multispectral satellite image is downloaded from Bhuvan, and this image is georeferenced in GCS WGS 1984 datum. Multispectral image is of size 1135 * 1135. This multispectral image belongs to Resourcesat-1 satellite. This image is having latitude of 6.75° and longitude of 93.5°. The characteristics of Resourcesat-1 satellite are mentioned in Table 1. The sensor used is Linear Imaging Self Scanning Sensor (LISS III). The image is visible in four bands i.e., green (G), infrared (IR), near infrared (NIR), and short wave infrared (SWIR). The spectral bands ranges for G, IR, NIR, and SWIR are (0.52–0.59) µm, (0.62–0.68) µm, (0.77–0.86) µm, and (1.55–1.7) µm, respectively. The four bands of multispectral image are fused to get false color composite (FCC) as shown in Fig. 2. This is done by composite band tool in ArcGis. The FCC image is then classified by interactive supervised classification as shown in Fig. 3. This technique does not require signature file. Maximum likelihood classification failed to classify the raster as there are huge numbers of pixels. Areas for three land covers are calculated as shown in Table 2. In this technique, some of water pixels are classified as land pixels. Due to this effect, the accuracies have been reduced shown in Table 4. Kappa coefficient and F1score are reduced as shown in Table 5 (Fig. 1 and Table 3).

In the proposed method, satellite image is given as input as shown in Fig. 1. Remote sensing indices NDVI, NDWI, and BSI are calculated by raster calculator in ArcGIS 10.3. NDVI, NDWI, and BSI maximum values are 0.778, 0.717, and 0.506, respectively. NDVI, NDWI, and BSI minimum values are 0.506, −0.613, and −0.089, respectively. NDVI, NDWI, and BSI raster's are shown Figs. 4, 5, and 6.

Table 1 Resourcesat/LISS main characteristics

Date of acquiring the image	22-3-2012
Projection	Geographic Lat/Long
Datum	WGS 84
Spatial resolution	22.5 m
Image file format	Geo-Tiff
Number of bands	4 (Band2, 3, 4, 5)
Radiometric resolution	8 bits
Temporal resolution	24 days
Swath	141 km

Table 2 Area calculation for land covers before feature enhancement

Land cover	Count	Area (km^2)
Land	125,225	63.395
Water	1191	0.594
Vegetation	2851	1.443

Table 3 Area calculation for land covers after feature enhancement

Land cover	Count	Area (km^2)
Land	107,959	54.654
Water	1385	0.701
Vegetation	3282	1.661

Table 4 Error matrix before feature enhancement

Land cover	Land	Vegetation	Water	Row total	Users accuracy (%)	Producers accuracy (%)
Land	**24**	0	24	48	50	96
Vegetation	0	**17**	1	18	94.44	85
Water	1	3	**1**	5	20	3.85
Column total	25	20	26	71		
Overall accuracy	57.53%					
Overall kappa coefficient	0.386					

Bold numbers are true value of pixels in each class which is classified

Table 5 Parameters before feature enhancement

Land cover	Commission (%)	Omission (%)	Kappa coefficient	F1score
Land	50	20	0.463	0.657
Vegetation	5.555	15	0.560	0.894
Water	80	3.846	0.580	0.064

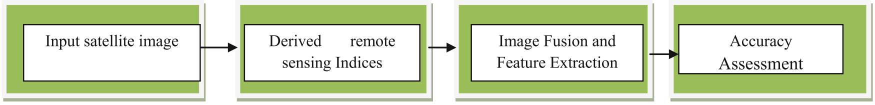


Fig. 1 Block diagram of the proposed method

Fig. 2 FCC for bands 4-3-2

Fig. 3 Classified image

Fig. 4 NDVI image

The area for three land covers is calculated shown in Table 3. All these vegetation indices raster's are combined or fused by composite bands tool in ArcGIS as shown in Fig. 7. The image features are enhanced after feature extraction. For accuracy assessment, image is classified using interactive supervised classification as shown in Fig. 8. Interactive supervised classification does not need signature file. Here, maximum likelihood classification failed to create signature file as there are more number of pixels. Ground truth data is generally obtained by field visit only, but here we can generate ground truth data by error matrix. The ground truth data is nothing but latitudes and longitudes of particular area. They are basically GPS points on earth. They are obtained by creating shape file in ArcGIS 10.3. More number of ground truth points on map is selected for better accuracy. Here, 28 points are selected in each land cover. The shape file and classified image is combined to get the combined raster. Finally, accuracy assessment is done for the three land covers, i.e., land, vegetation, and water. Producer's accuracy, user's accuracy commission, omission, kappa coefficient, F1score, overall accuracy, and overall kappa coefficients are calculated as shown in Tables 6 and 7. The proposed method has got better values compared to existing methods. Without fusion of remote sensing indices, raster's water bodies are not highlighted as shown in Fig. 3. Water has got least accuracies, kappa coefficient, F1score as shown in Tables 2 and 4. It has got more commission and omission, which is undesirable. Model calculations are shown from Eqs. 1 to 12.

Model calculations:

$$\text{Matching total} = \text{TP}_\text{A} + \text{TP}_\text{B} + \text{TP}_\text{C} \tag{1}$$

$$\text{TP} = \text{True positive, TN} = \text{True negative} \tag{2}$$

$$\text{Over all Accuracy} = \frac{\text{Matching Total}}{N} \tag{3}$$

$$\text{Matching Pixel} = \text{Diagonal elements in error matrix}$$

Fig. 5 NDWI image

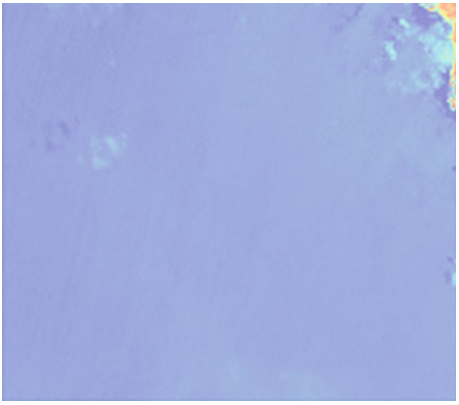


Fig. 6 BSI image

$$\text{Producers accuracy} = \frac{\text{Matching Pixels}}{\text{Reference pixels}} \tag{4}$$

$$\text{Reference Pixels} = \text{column wise}$$

$$\text{Users accuracy} = \frac{\text{Matching Pixels}}{\text{Classified Pixels}} \tag{5}$$

$$\text{Classified pixels} = \text{row wise}$$

Fig. 7 FCC after NDVI, NDWI, and BSI fusion

Fig. 8 Classified image after fusion

$$\text{Commission} = \frac{\text{Incorrectly classified pixels in row (TN Total)}}{\text{Total No. of pixels of the row}} \tag{6}$$

$$\text{Omission} = \frac{\text{Incorrectly classified pixels in column (TN Total)}}{\text{Total No. of pixels of the column}} \tag{7}$$

$$\text{Kappa coefficient} = N * \frac{\sum_{i=1}^{r} X_{ii} - \sum_{1=1}^{r} (X_{i+} * X_{+i})}{\left(N^2 - \sum_{i=1}^{r} X_{i+} * X_{+i}\right)} \tag{8}$$

N total number of pixels
X_{ii} sum of correct pixels
X_{i+} sum of all row total
X_{+i} sum of all column total

Table 6 Error matrix after feature enhancement

Land cover	Land	Vegetation	Water	Row total	Users accuracy (%)	Producers accuracy (%)
Land	**24**	1	9	34	70.58	96
Vegetation	0	**21**	1	22	95.45	87.5
Water	1	2	**17**	20	85	62.96
Column total	25	24	27	76		
Overall accuracy	81.58%					
Overall kappa coefficient	0.724					

Bold numbers are true value of pixels in each class which is classified

Table 7 Parameters after feature enhancement

Land cover	Commission (%)	Omission (%)	Kappa coefficient	F1score
Land	29.411	4	0.784	0.814
Vegetation	4.545	12.5	0.798	0.913
Water	15	37.037	0.797	0.723

$$\text{F1score} = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \tag{9}$$

$$\text{NDVI} = \frac{\text{NIR} - \text{IR}}{\text{NIR} + \text{IR}} \tag{10}$$

$$\text{NDWI} = \frac{\text{G} - \text{NIR}}{\text{G} + \text{NIR}} \tag{11}$$

$$\text{BSI} = \frac{\text{SWIR} + \text{IR} - \text{NIR}}{\text{SWIR} + \text{IR} + \text{NIR}} \tag{12}$$

3 Conclusion

Remote sensing indices are calculated for multispectral satellite image. NDVI, NDWI, and BSI raster's are combined. By combining this raster's the features of image has been enhanced. The resultant satellite image is classified into three land covers, i.e., water, vegetation, and land. Error matrix is used to generate ground truth data. Then, quality parameters are calculated for the particular classification technique. Producer's accuracy and user's accuracy, kappa coefficient, commission, omission, F1score has been calculated. The proposed method has got better values compared to the existing technique.

Acknowledgements Authors are grateful to Bhuvan for providing satellite images to our research work.

References

1. Gonzalez RC, Woods RE (2008) Digital image processing. Prentice Hall, New Jersey
2. Lee JS, Grunes MR, Schuler DL, Pottier E, Ferro-Famil L (2006) Scattering-model based speckle filtering polar metric SAR data. IEEE Trans Geosci Remote Sens 44:176–187
3. Lillesand TM, Kiffer RW (2000) Remote sensing and image interpretation, 4th edn. Wiley, New York
4. Meddens AJ, Hicke JA, Vierling LA, Hudak AT (2013) Evaluating methods to detect bark beetle-caused tree mortality using single-date and multi-date Landsat imagery. Remote Sens Environ 132:49–58
5. Willis KS (2015) Remote sensing change detection for ecological monitoring in United States protected areas. Biol Conserv 182:233–242
6. Randon J, Hüsoy JH (1999) Filtering for texture classification: a comparative study. IEEE Trans Pattern Anal Mach Intell 21:291–310
7. Justice CO, Vermote E, Townshend JRG et al (1998) The Moderate Resolution Imaging Spectro radiometer (MODIS): land remote sensing for global change research. IEEE Trans Geosci Remote Sens 36:1228–1249
8. Hudson WD, Ramm CW (1987) Correct formulation of the kappa coefficient of agreement. Photogram Eng Remote Sens 53:421–422
9. Rokni K, Ahmad A, Selamat A, Hazini S (2014) Water feature extraction and change detection using multitemporal Landsat imagery. Remote Sens 6:4173–4189
10. Molchanov V, Chitiboi T, Linsen L (2015) Visual analysis of medical image segmentation feature space for interactive classification. In: Eurographics conference, pp 11–19

Detection of Heart Abnormalities and High-Level Cholesterol Through Iris

P. A. Reshma, K. V. Divya and T. B. Subair

Abstract Iridology is medicine technique to claim the colours, characteristics and pattern of the iris. Iridology is used to determine the existence of basic genetics, irregularities in the body, dam circulation, toxin deposition and other weakness. This paper discusses the determination of high-level cholesterol in blood and heart conditions through several stages. In this study, we examine the heart condition and cholesterol through preprocessing, segmentation, feature extraction and classification from the captured iris image. Due to the high level of cholesterol in blood, sodium ring is to be formed around the iris.

1 Introduction

The heart is an internal organ of our body. This organ provides blood supply to our body by coronary arteries. Heart disease is known as a cardiovascular disease. Now today, heart problems are increasing day by day. A number of factors that cause heart disease are smoking, high blood pressure, high level of cholesterol, diabetes, obesity or from the family history. Different symptoms and signs are occurring in heart abnormalities. One of the methods is iridology. According to iridology, certain areas of human iris represent a particular organ condition [1]. Figure 1 shows a particular organ. Iris of the eye is used to detect the condition of a body organ, genetic strength and weakness. This method believes that patterns on the iris reflect body condition [2].

P. A. Reshma (✉) · K. V. Divya
Department of Computer Science Engineering, Vidya Academy of Science and Technology, VAST, Thissur, Kerala, India
e-mail: reshmarazack01@gmail.com

K. V. Divya
e-mail: divyakv@vidyaacademy.ac.in

T. B. Subair
Technician-Classroom Technologies, Information Technology Centre, University of Sharjah, Sharjah, UAE
e-mail: sbeeravu@sharjah.ac.ae

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_35

Fig. 1 Structure of a human eye with particular organ condition

Iris, in communication between brain and organs, has approximately 28,000 nerve network. That is, an organ irregular tries to send information about the state of the brain and this information on the pattern of the iris, as reflected in a change in the colour or characteristic [3]. Figure 1 shows some examples of signs of organ and tissue disorder.

Cholesterol is an oily substance that will not mix with the blood. This substance plant in our body cells. Our body contains two kinds of cholesterol: LDL and HDL. Both types are decisive for our body health. Blood cholesterol levels are measured by subtracting the lipoprotein profile, 9–12-h measurement of the total cholesterol level in the blood after starving with cord blood. Generally, LDL is a bad cholesterol and HDL is a good cholesterol. High cholesterol level causes heart abnormalities, diabetes, PCOS and kidney disease. In our body, the level of cholesterol increases on consuming cheese, animal foods, meats, baked goods, chocolate, deep-fried goods, etc.

The iris pattern has many distinguishing metabolism characteristics in the human body as well as there is also the ability to outplay changes [4]. Major studies with iridology between A. D. Wibawa and M. H. Purnomo diagnosis of mellitus diabetes by looking at the status of the pancreas [5], Cheng-Liang Lai and Chien-Lun Chiu's heart [6].

In general, blood cholesterol levels are measured by subtracting the lipoprotein profile. Lipoprotein profile, 9–12-h measurement of the total cholesterol level in the blood after starving with cord blood [7]. Blood cholesterol level increases light-coloured layer of sodium that begins to occur around the outer ring of iris. Increased level of cholesterol can detect with the iridology property [8]. This research will be able to contribute to the detection of high-level cholesterol and disorders of the heart through the iris image. This helps to give the information about our health disease.

2 Detection of Heart Abnormalities

The heart abnormalities detection [9] system simply illustrates that the input image is trained by preprocessing and feature extraction process. In a preprocessing step, the acquired image tests. After the feature extraction step, we assign the thresholding value and then classify whether the patient is abnormal or not.

2.1 *Image Acquisition*

High-resolution and specification camera are used for acquiring the iris image, for accurately getting the features of iris. We use the camera to take the image with the following specifications:

- 12.8 Megapixel
- 28–224 mm—Lens Focal Length
- Digital Zoom 4x, Optical Zoom 10x
- 1.97″–19.69″—Infinity Macro
- Back-illuminated
- Additional flashlight

The additional flashlight is used for getting the features more accurately. Figure 2 shows the difference in using an additional flashlight and not.

2.2 *Preprocessing*

The preprocessing stage is very important for iris recognition process. It increases the quality of the original image and determines the successful result. There are different noise removal filtering techniques available: removing the noise by linear filtering, averaging filter, median filter and an adaptive filter. In here, removing noise from the captured iris image by using a median filter.

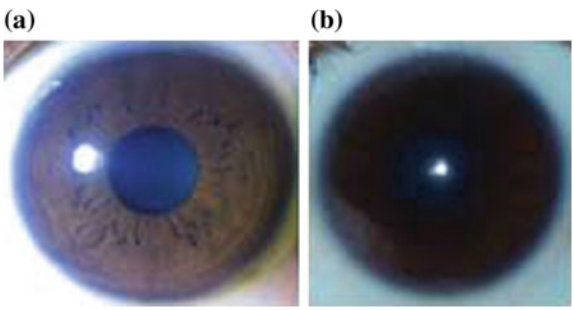


Fig. 2 Iris image with an additional flashlight and not

Fig. 3 Binarized iris image

In the median filter, the output pixel value is determined by the median of the neighbourhood pixels. The median filter is better to eliminate the outliers without compressing the sharpness of the image. Here, the calculation is done on the basis of the pixel intensity values (Fig. 3).

After the noise is removed, the image is cropped. Image cropping is done by binarization of image and histogram. The binarization process converts the iris image into a binary image. The binary image is like a white and black image. The value of the white pixel and a black pixel is divided into horizontals and verticals. In cropping process, the image can be categorised as success, fail and scant. The scant image produces less perfect iris image. The full boundary of the image cannot be gained. The failed image is an incomplete one. The causes of the scant image are shadowing and improper lighting.

2.3 *Feature Extraction*

From the result of segmentation, the step gives the information about the disease. It is calculated by the ratio of a number of white and black pixel.

$$\text{Ratio of Black} = \text{Total Black}/\text{Total Pixels} \quad (1)$$

$$\text{Ratio of White} = \text{Total White}/\text{Total Pixels} \quad (2)$$

2.4 *Classification*

The classification is done on the basis of thresholding value. The threshold value is determined by analysis of test data. The threshold value of black is 0.50913 and white is 0.48848. First, enter the data to be classified. Then the label (normal or abnormal) is produced if the entered data value is below or above the threshold value.

3 Detection of Cholesterol

In cholesterol detection [10] process, first images are acquired from high cholesterol patient and normal patient. Then the pupil of the iris is used to determine the width of sodium ring by analysing the colour tone. The cholesterol detection process is explained below.

3.1 Image Acquisition

In cholesterol detection process, first images are acquired from high cholesterol patient and normal person. Figure 4 shows the iris image of cholesterol patient, and Fig. 5 shows the normal person.

3.2 Localization

Here localization process means finding the pupil from the iris. The iris, pupil and eye are of different colours, rather than taking advantage of the Adam edge discovery process. For finding the boundaries of the pupil by integral differential operator. The IDO was proposed by John Daugman [11]. Daugman first proposed iris recognition.

Fig. 4 Iris image of high-level cholesterol patient

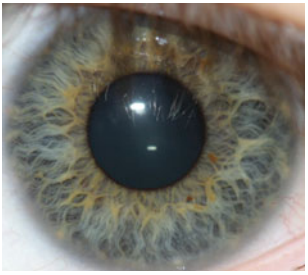


Fig. 5 Iris image of a normal person

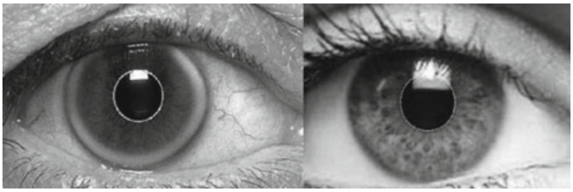


Fig. 6 Locate the pupil boundary

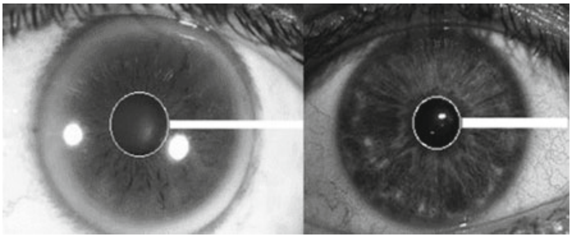


Fig. 7 Horizontal and vertical axis

3.3 Used Method

The IDO method determines the boundaries of the iris with a very high accuracy rate. The boundaries are calculated by finding the radius of pupil and radius of iris. The radius of inner and outer circles is in the range from 0.1 to 0.8 of the iris radius [12]. Equation 1 explains the IDO (integrodifferential operator).

$$\max(r, x_0, y_0)\left|G_\sigma(r) * \frac{\partial}{\partial_{r_{r,x_0,y_0}}} \oint \frac{-I(x, y)}{2\pi r} \mathrm{d}s\right| \tag{3}$$

where I (x, y) in the (x, y) position of a hue value, x_0 and y_0 is possible to centre coordinates, r distance to the point of the possible central symbol, $G_\sigma(r)$ is a smoothing function such as a standard offset Gaussian function. By using this method, the position of the eye can be found correctly. The reflections in scanning process of the picture are ignored. Reflection of the boundaries of the eyeball is unaffected. Figure 6 simply explains it.

Sodium ring thickness is calculated by the ratio between the width of the iris and width of the sodium ring. This will be examined by taking the pupil centre as a vertical axis, and the horizontal axis starts from the right side of the eye continues to the end of the image. Figure 7 shows the pictorial representation.

Table 1 shows the sodium ring eye thickness ratio results. Picture number 4, the rate at which the database with a value of sodium ring 0.09 thinnest sodium ringed eye. The eye number five with a value of 0.23% sodium ring is the boldest look (Fig. 8).

Table 1 The ratio of the thickness

Resim Numarasi	Oran
1	0.13
2	0.19
3	0.11
4	0.09
5	0.23
6	0.16

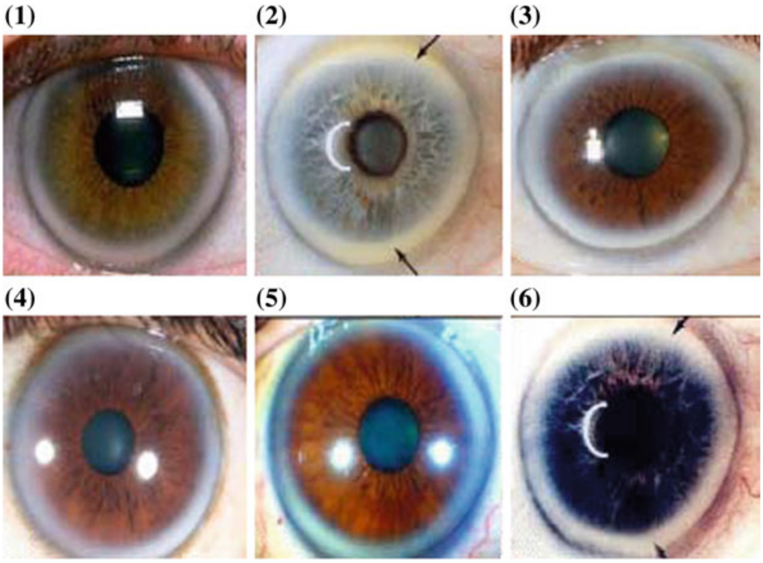


Fig. 8 Thickness of sodium ring

4 Conclusion

One of the most popular treatments is iridology. The iris image is high quality and perfectly add flashlight then the result will be accurate. The cropping process is determining the comparison of the threshold value. The ratio between the width of the iris with the width of sodium ring is computed, and evaluations have been conducted. The rate at which values are obtained as a result of evaluations for each eye is different. Therefore, iris sodium ring especially high cholesterol value is a direct relationship between can be associated. The cholesterol disease and heart abnormality are closely related, so that this research has a huge potential for future.

References

1. Hiru (2005) Iridology: Mendeteksi Penyakit Hanya dengan Mengintip Mata. Gramedia Pustaka Utama
2. Jensen B (2005) Science and practice of iridology
3. Woodward JD, Orlans NM, Higgins PT (2003) Biometrics: identity assurance in the information age. ISBN 0-07-222227-1
4. Sivasankar K, Sujaritha M, Pasupathi P, Muthukumar S (2012) FCM based iris image analysis for tissue imbalance stage identification. In: International conference on emerging trends in science
5. Wibawa D, Purnomo MH (2006) Early detection on the condition of pancreas organ as the cause of diabetes mellitus by real time iris image processing. In: IEEE Asia Pacific conference on circuits and systems
6. Lai C, Chiu C (2010) Health examination based on iris images. In: Proceedings of the ninth international conference on machine learning and cybernetics, Qingdao
7. Ramlee RA, Ranjit S (2009) Using iris recognition algorithm, detecting cholesterol presence. In: International conference on information management and engineering
8. NutritionalIridology.com (2012) Iridology charts. Retrieved 17 Oct 2013, from Nutritional Iridology. http://nutritionaliridology.com/charts.html
9. Entin Martiana K, Ridho A, Syarifa S, Afgan A (2016) Application for heart abnormalities detection through iris. IES, IEEE. 978-1-5090-1640-2/16
10. Kursat B, Kurnaz C (2016) Detection of high-level cholesterol in blood with iris analysis. IEEE. 978-1-5090-5829-7/16
11. Daugman J (2009) Iris recognition at airports and border-crossings. In: Encyclopedia of biometrics, pp 819–825
12. Daugman J (2004) How iris recognition works. IEEE Trans Circ Syst Video Technol 14(1):21–30

Wavelet-Based Convolutional Recurrent Neural Network for the Automatic Detection of Absence Seizure

Kamal Basha Niha and Wahab Aisha Banu

Abstract In this paper, the new model is proposed to automatically detect and predict absence seizure using hybrid deep learning algorithm [Convolutional Recurrent Neural Network (CRNN)] along with the Discrete Wavelet Transform (DWT) with Electroencephalography (EEG) as input. This model comprises of four steps (1) Single-channel segmentation process (2) Decomposition of segmented signal using wavelet transform (3) Extraction of relevant feature using statistical method (4) Deep learning algorithms for classification, detection, and early detection. This model enhances the feature extraction and also the overall performance by feeding the segmented data into Long Short Tern Memory (LSTM) which is one of the Recurrent Neural Network (RNN). And also the output of this network is used to calculate the extracted feature along with the classification results. The values in hidden state are used to diagnose the seizure by locating the pattern using the extracted features of time window. The proposed model achieves 100% accuracy on detection and 95% overall accuracy on early detection of normal, abnormal and absence seizure.

1 Introduction

Absence seizure is one among the types of epilepsy where its prevalence is about one percent among the population which means eight point eight per thousand population and it is higher in the rural area (one point nine percent) than urban (zero point six percent). Absence seizures occurrence lasts from short to long duration in terms of minutes. In general, its spike and wave pattern of frequency fall between

K. B. Niha (✉) · W. Aisha Banu
Computer Science and Engineering, B S Abdur Rahman Crescent Institute of Science and Technology, Chennai, India
e-mail: niha.k.cse@gmail.com

W. Aisha Banu
e-mail: aisha@crescent.education

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_36

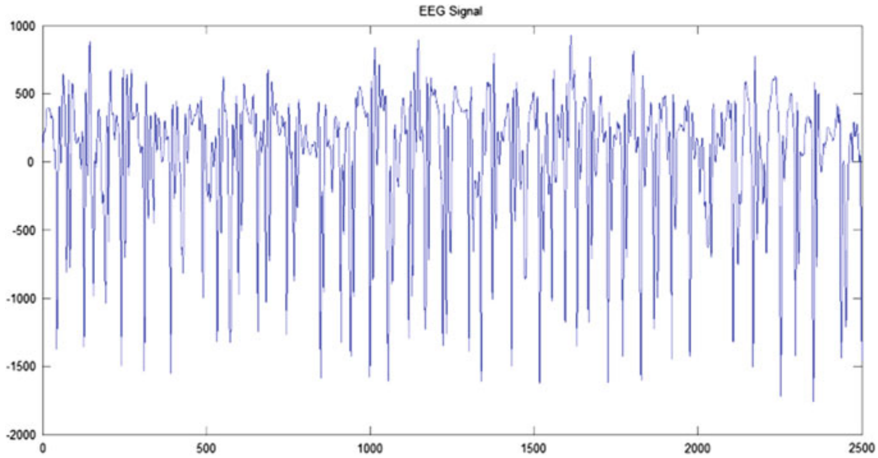


Fig. 1 Single-channel raw absence seizure EEG data

2 and 3 Hz per ms (mille second) and in terms of duration it is 5–30 s long. This patter is found in the frontal region, which may start around 4 per sec and slows down to 3–3.5 per sec then at last it decreases to 2.5 per sec. The single-channel raw seizure EEG data are shown in Fig. 1. In traditional medicine, seizures frequency has been computed manually for analysis and diagnosis are in practice. And also monitoring patients out of clinical boundary with the help of portable EEG recording systems, the same frequency seizure signals are able to obtain for in-patients are in practice. But this process is very tedious and leads to many measurement errors. Only an experienced physician can able to handle this diagnosis process. To overcome this, seizures are diagnosed by automatic detection and early detection systems [1] which has been designed using machine learning algorithms.

Among those algorithms, Conventional Neural network (CNN) and Recurrent Neural Network (RNN) play a major role. Initially, CNN was used widely in different fields for classification of input in computer vision, recognition of traffic signal, and generating image captions [2–4]. This algorithm is not only restricted to the above fields and also for signal processing domain [5]. Major advantage of this algorithm is that it outperforms the handcrafted feature extraction process [6]. In parallel, the RNN also in boom which is widely used to handle the sequence of data with varied sequence. These are most successful in speech recognition and machine translation [7–9]. When the input is in sequence with varied length, both CNN and RNN can be used. With that features are extracted from segments of data and those extracted feature form another one sequence. These sequences are taken into recurrent layer with temporal dependency can be handled in CNN using pooling mechanism. The main disadvantage in traditional CNN is to extract features they use nonlinear affine function. The advantage of this work by using CRNN is to compute the better features when compared to the traditional convolutional layer. The upcoming section briefly discusses experimental design, working method, results, and conclusion.

2 Experimental Design

The processing flow of this model is shown in Fig. 2. In that, the subjects EEG signal are recorded using clinical setup and are viewed using RMS software. Then, those signals are processed with wavelet transform and CRNN (LSTM) to obtain result.

Initially, the scalp EEG data have been recorded using clinical setup (16 channel electrodes) based on 10–20 international system. The duration of absence seizure data will last for 2 s. The sampling frequency of 512 Hz and the bandpass filter from 1.0 to 70 Hz have been used during the recording and to the power line noise notch filter have been applied as 50 Hz to remove power line noise. Duration of each recorded data is for 24 s. These recorded data are viewed using RMS software and from that single-channel (FP1-F7) data which is more reliable and sensitive [9] than other channel. Then, the data have been exported to MATLAB environment.

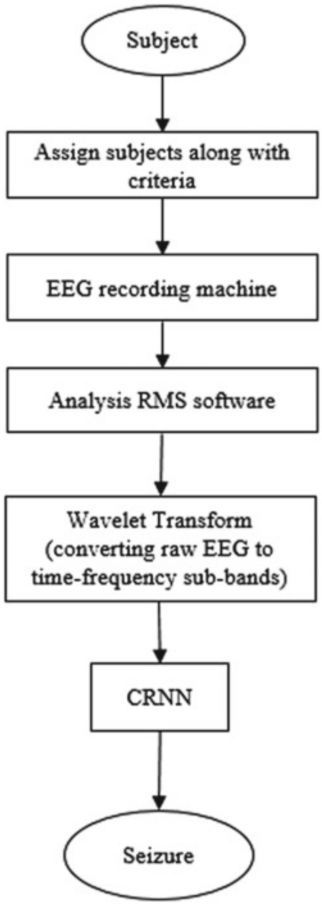


Fig. 2 Work flow of the CRNN system

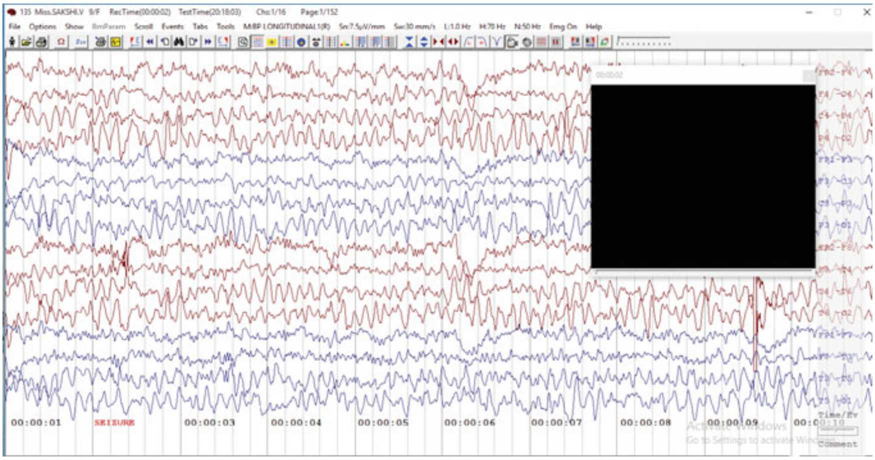


Fig. 3 RMS View of absence seizure EEG data with spike and wave

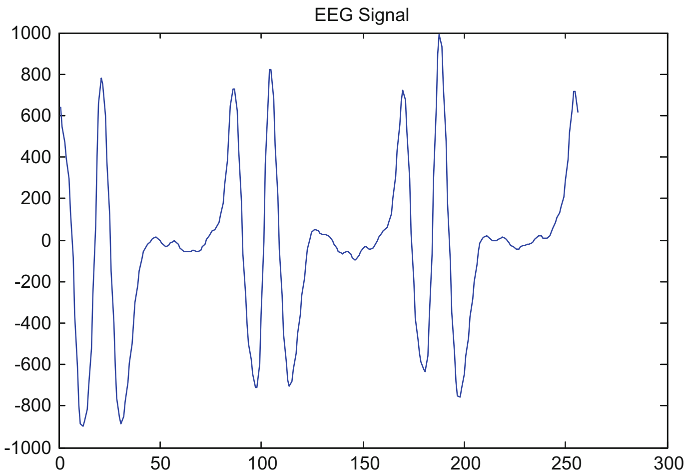


Fig. 4 Single-channel segmented absence seizure EEG data

The RMS view of absence seizure data and the segmented data have been shown in Figs. 3 and 4. The sequence of data has been segmented as 258 samples each and then these segments are undergone for time-frequency transformation using Discrete Wavelet Transform [10].

The transformed output contains five-time domain and frequency domain sub-bands. They are alpha, beta, gamma, delta, and theta. Their frequency ranges are

- Delta (0.1–3.5 Hz), Theta (4–7.5 Hz)
- Alpha (8–13 Hz), Beta (14–40 Hz)
- Gamma (greater than 40 Hz).

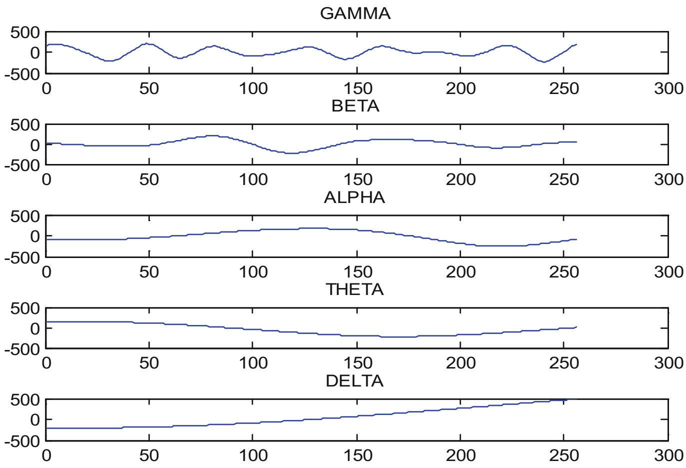


Fig. 5 Time domain absence seizure EEG data

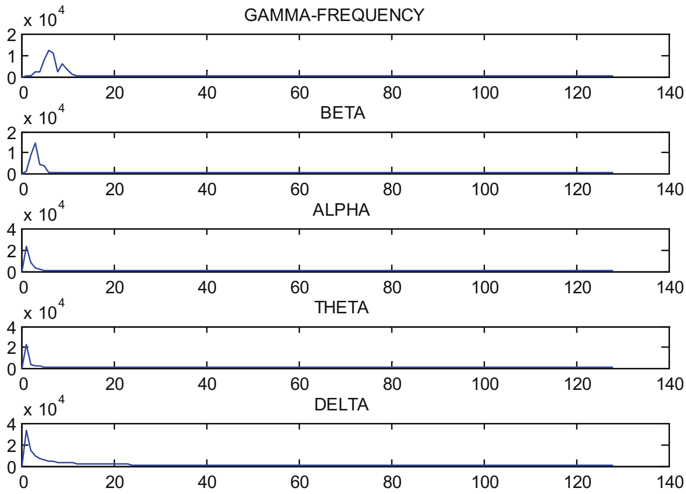


Fig. 6 Frequency domain absence seizure EEG data

These sub-bands are shown in Figs. 5 and 6. Each sub-band is processed using Convolutional Recurrent Neural Network (CRNN) for detection and early detection of seizure. Where the feature extractions are handled automatically and are classified to diagnose the seizure. Long Short-Term Memory (LSTM) has been used for early detection.

3 Method

The proposed Conversional Recurrent Neural Network in this paper extracts the EEG features by itself from the sub-bands of time domain. Initially, the absence seizure dataset comprises of varied time series of data, each sequence consists of frames and each frame has the same fixed number of features. Here, p is the sequence of data, q is the frame consist of feature vector with length r, and the features are represented as r. In this, the product of the size of the feature vector and the frame (q * r) is equal to the sequence of data p. The sequence of input window p comprises of $s_1, s_2, \ldots, s_n$ where the window size is equal to the product of r and s_1. Then apply the function f which extracts n features from each window which generates sequences of x' which comprises of n features with respect to each frame. After this step, the pooling mechanism has been applied to pick the maximum value among the extracted feature of each window of size $n * m_1$ from x' which comprises of $m_1, m_2, \ldots, m_n$ frames. After max pooling, the size of the frame gets transferred to windows of size $n * m_1$ to the vector of size $n * 1$. Finally, the local features of sequence p have been feed into the next layer for the classification. Where LSTM has been used as a recurrent layer in this model which generates the additional sequences they are output ($y_1, y_2, \ldots, y_{r1}$) and cell state ($c_1, c_2, \ldots, c_{r1}$) instead of hidden state to compute the features representing the windows. This model works on the 3 class classification problem (normal, ictal and absence ictal). The LSTM consists of 128 hidden layer followed by 100 hidden nodes with fully connected (f_c) layer. This layer has been applied on the continuous sequence with time distributed manner. Here the linear activation function has been used and the output is given to another 128 LSTM hidden layer. The class probability has been computed using the output of the second LSTM layer. For this, the schematics of LSTM are given below,

Input layer_1
LSTM_0
Time distribution_1
LSTM_1
Dens_2

The characteristics of model are given below as parameters and its value, hidden unit layer_1 as 100, hidden unit layer_2 as 100, fully connected layer_2 as 100, gate activation as sigmoid, cost as cross-entropy (categorical), epoch as 200, optimizer as Adam, dropout as 40, batch size as 25. With these characteristics, the model has been trained in state full mode in which the data stored in the previous layer of internal memory propagate among samples. Where the internal memory of the LSTM has been initialized to 0 and the training has been done with Adam stochastic optimizer. The learning rate of this model is 0.001 and the model is trained using 200 epoch and validated using k-fold cross-validation.

Table 1 Confusion matrix of the LSTM

	Normal	Abnormal	Absence
Normal	12	3	0
Abnormal	0	15	0
Absence	0	0	8

4 Result

The classification accuracy has been analyzed using the performance metrics. The high accuracy classification cannot have high prediction accuracy.

To cross check that in this paper, the performance of this model has been evaluated using precision, recall, and F_1 measures. To calculate these metrics, confusion matrix (Table 1) has been created for testing data. To compute accuracy, precision, recall, and F1 score formulas are given below

$$\text{Precision} = T_p/T_p + F_p$$
$$\text{Recall} = T_p/T_p + T_n$$
$$\text{Accuracy} = T_p + T_n/T_p + F_p + T_n + F_n$$
$$F_1 = 2 * ((\text{Precision} * \text{Recall})/(\text{Precision} + \text{Recall}))$$

Calculated precision, recall, accuracy, and F_1 measures of the model are 95, 94, 95, and 94%.

5 Conclusion

In this paper, the novel RNN layer has been used for EEG absence seizure classification. Here, the RNN refers to LSTM which have been combined together with CNN to obtain a hybrid model with the accuracy of 95%. This model is very easy to train because there is no need of feature extraction and selection process. This model is less computational for fast training and better for large dataset. In future, different combinations of deep learning algorithms are used for better classification accuracy.

Acknowledgements We thank Dr. S. Velusamy, DM—Neurology, MD—Paediatrics, MBBS Neurologist, and General Physician, who has 22 years of experience for his continuous support throughout this work.

References

1. Schelter B, Winterhalder M, Maiwald T, Brandt A, Schad A, Timmer J, Schulze Bonhage A (2006) Do false predictions of seizures depend on the state of vigilance? A report from two seizure-Prediction methods and proposed remedies. Epilepsia 47(12):2058–2070

2. Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural networks. In: Proceedings of advances in neural information processing systems (NIPS), Lake Tahoe, NV, pp 1097–1105
3. Ciresan D, Meier U, Schmidhuber J (2012) Multi-column deep neural networks for image classification. In: Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR), Providence, RI, pp 3642–3649
4. Vinyals O, Toshev A, Bengio S, Erhan D (2015) Show and tell: a neural image caption generator. In: Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR), Boston, MA, pp 3156–3164
5. Sainath TN, Weiss RJ, Senior A, Wilson KW, Vinyals O (2015) Learning the speech front-end with raw waveform CLDNNs. In: Proceedings of INTERSPEECH 2015, 16th Annual conference of the international speech communication association, Dresden, Germany, p 15
6. Amodei D, Anubhai R, Battenberg E, Case C, Casper J, Catanzaro BC, Chen J, Chrzanowski M, Coates A, Diamos G, Elsen E, Engel J, Fan L, Fougner C, Han T, Hannun AY, Jun B, LeGresley P, Lin L, Narang S, Ng AY, Ozair S, Prenger R, Raiman J, Satheesh S, Seetapun D, Sengupta S, Wang Y, Wang Z, Wang C, Xiao B, Yogatama D, Zhan J, Zhu Z (2015) Deep speech 2: end-to-end speech recognition in English and Mandarin. arXiv, preprint arXiv:1512.02595
7. Graves A, Mohamed A, Hinton GE (2013) Speech recognition with deep recurrent neural networks. In: Proceedings of the IEEE international conference on acoustics, speech and signal processing, (ICASSP), Vancouver, Canada, pp 6645–6649
8. Cho K, Van Merrienboer B, Gulcehre C, Bahdanau D, Bougares F, Schwenk H, Bengio Y (2014) Learning phrase representations using RNN encoder–decoder for statistical machine translation. In: Proceedings of the conference on empirical methods in natural language processing (EMNLP), Doha, Qatar, pp 1724–1734
9. Sutskever I, Vinyals O, Le QV (2014) Sequence to sequence learning with neural networks. In: Proceedings of advances in neural information processing systems (NIPS), Montreal, Canada, pp 3104–3112
10. Zhang B, Jiang H, Dong L (2017) Classification of EEG signal by WT-CNN model in emotion recognition system. In: International conference on Informatics and cognitive computing. IEEE

New Random Noise Denoising Method for Biomedical Image Processing Applications

G. Sasibhushana Rao, G. Vimala Kumari and B. Prabhakara Rao

Abstract Since the inception of digital image processing, noise removal in images has always been a challenge to researchers and experts of the field. Most significant of these noises are the randomly varying impulse noises developed while image is acquired. Hence, the need for methodical denoising method has led to extensive research and development of various innovative methods to remove the random valued impulse noise. For this, a method which detects and filters random valued impulse noise in medical images is employed. The method proposed in this paper uses a decision tree based impulse detector and an edge preserving filter to rebuild noise free images. This method is more efficient than the existing techniques due to its lower complexity. Different gray scale Magnetic Resonance Imaging (MRI) brain images are tested by using this algorithm and have given better Peak Signal to Noise Ratio (PSNR) than the other techniques.

1 Introduction

With the advent of digital images, productivity in science and technology has greatly increased due to better understanding capabilities, analysis, visualization and interpretation; which are not limited to a particular field but can be extended to as many

G. Sasibhushana Rao
Department of Electronics and Communication Engineering, AU College of Engineering, Visakhapatnam 530003, India

G. Vimala Kumari (✉)
Department of Electronics and Communication Engineering, M.V.G.R College of Engineering, Vizianagaram 535002, India
e-mail: Vimalakumari7@gmail.com

B. Prabhakara Rao
Department of Electronics and Communication Engineering, JNTUK, Kakinda 533003, Andhra Pradesh, India

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_37

as possible. However, with these advantages came the limitations of using digital images. Prominent among these include removing noise and storage problems. Noise develops in images while the image is acquired and transmitted. Presence of noise in images cripples all the advantages with using digital images as it deteriorates the quality of images making it difficult to analyse, interpret or visualize [1, 2]. This led to the development of effective denoising algorithms which target various types of noises in images. Among the discovered image noises, most arduous are the impulse noises which affect the images during their acquisition and transmission [3]. Impulse noises generally involve random occurrence or distribution of noisy pixels over the image. However, based on the values of its noisy pixels, these are divided into two categories as fixed value impulse noise and random-valued impulse noise. Fixed value impulse noise is limited to only two noisy pixel values, i.e. salt—255 and pepper—0. On the other hand, random-valued impulse noise involves noisy pixels of any value within the range [0–255] [4, 5]. Thus, there exists a need for development of an efficient denoising method for removal of random valued impulse noise in images. Mean filters perform denoising effectively, but at the cost of heavy information loss due to blurring or smoothening of the image. Median filters are preferred to preserve several important details even after denoising. Although not as much as mean filter, median filters result in information loss too [6, 7]. Hence, to avoid this switching median concept is opted. Here, the median filter is provided with an impulse detector which detects the noisy pixels prior to the filter. Thus by detecting the noisy pixels before filtering, the filter may be conditioned such that it filters only those noisy pixels that are detected thereby resulting in less information loss and better preservation of details in the image [8, 9]. Adaptive Median Filter raises the size of the processing kernel to get rid of noisy pixels. Further this filter offers best results in getting rid of noisy pixels at lower densities. Yet, the limitation of this filter is that it blurs the image at higher densities. To overcome this limitation the Adaptive Decision Based Median Filter (ADMF) [10] is proposed. A Nonlocal Means (NLM) based filtering algorithm for denoising Rician distributed MRI and utilizes the concept of self-similarity for MRI restoration [11]. An approach involving minimum absolute difference criteria for identifying the noisy pixels and replacement by mean value computed [12]. It is evident that denoising random-valued impulse noise is far more challenging than denoising salt-and-pepper noise due to its random pixel values. Of all the various methods available to denoise impulse noise, very few address the random-valued impulse noise. Thus, there exists a need for the development of an efficient denoising method for removal of random-valued impulse noise in images. Keeping in view the above requirements, in this paper a model which is of low complexity that employs a new impulse detector based on decision tree and a filter to preserve the edges to denoise random-valued impulse noise in an image is proposed.

2 Proposed Algorithm

The algorithm proposed in this paper uses a modified Decision Tree Based Denoising Method (DTBDM) to reconstruct noise-free images. DTBDM method involves two main components, viz., impulse detector and edge preserving filter. The impulse detector makes use of a 3 × 3 mask, to identify the noisy pixels and to activate the filter accordingly. The centre pixel is denoted by $p_{i,j}$, while its luminance is denoted by $s_{i,j}$. The neighbours are named accordingly in the range of ($i - 1$: $i + 1$, $j - 1$: $j + 1$).

2.1 Impulse Detector

The impulse detector detects the corrupted pixels so as to activate the filter only for those pixels that are corrupted. Here, an impulse detector is designed based on a decision tree. The impulse detector comprises three modules, viz., segregation, perimeter and analogous. Each of these modules is independent of each other and can segregate corrupted pixels. The pixel is considered noise-free only if all three modules judge it to be uncorrupted. Several thresholds, viz., Th_1, Th_2, Th_3, Th_4, Th_5 and Th_6 are used at various places in the modules. The values of Th_1, Th_2, Th_3, Th_4, Th_5 and Th_6 are 15, 20, 45, 70, 10 and 50, respectively.

Segregation Module. Segregation logic is based on the prime assumption that the centre pixel is located on a smooth surface. Thus, to test the pixel for segregation, first, it has to be on a uniform surface. Practically, the impulse module can be easily implemented using the 3 × 3 mask. The mask is first split into two halves $U_{T\text{-Half}}$ and $U_{B\text{-Half}}$. Now, the nine pixels in the mask can be named as 'a, b, c, d, e, f, g, h and $s_{i,j}$', where $s_{i,j}$ denotes the centre pixel and the variables 'a to h' denote the eight neighbours of the centre pixel. $U_{T\text{-Half}}$ comprises the top four neighbours 'a to d', while, $U_{B\text{-Half}}$ comprises the four neighbours 'e to h', at the bottom of the mask. They are represented as

$$U_{T-\text{Half}} = \{a, b, c, d\} \tag{1}$$

$$U_{B-\text{Half}} = \{e, f, g, h\} \tag{2}$$

The test for uniformity can be performed by finding out the maximum possible difference in intensities of both the top and bottom halves of the mask. If either of these differences in intensity is very high, i.e. greater than a threshold, then the region is considered as a nonuniform region and the test for segregation cannot be performed. The test can be represented using the following equations

$$U_{T-\text{Halfdif}} = U_{T-\text{Halfmax}} - U_{T-\text{Halfmin}} \tag{3}$$

$$U_{B-\text{Halfdif}} = U_{B-\text{Halfmax}} - U_{B-\text{Halfmin}} \tag{4}$$

$$\text{Decision 1} = \begin{cases} \text{true}, & \text{if } (U_{T-\text{Halfdif}} \geq \text{Th\_1}) or (U_{B-\text{Halfdif}} \geq \text{Th\_1}) \\ \text{false}, & \text{otherwise} \end{cases} \tag{5}$$

If the region is found to be uniform then proceed to the test for segregation. The centre pixel is now taken into consideration and the intensity difference between the centre pixel and $U_{T\text{-Halfmax}}$, $U_{T\text{-Halfmin}}$, $U_{B\text{-Halfmax}}$ and $U_{B\text{-Halfmin},}$ are computed, respectively. If any of these differences is higher than the threshold, then, the centre pixel is considered as an isolated pixel. Decision 2 determines whether the centre pixel is isolated or not. It is to be noted that Decision 2 is arrived at only when Decision 1 is false. Isolated pixel can be represented as

$$I_{T-\text{Half}} = \begin{cases} \text{true}, & \text{if}\left(|S_{i,j} - U_{T-\text{Halfmax}}| \geq \text{Th\_2}\right) \\ \text{or } \left(|S_{i,j} - U_{T-\text{Halfmin}}| \geq \text{Th\_2}\right) \\ \text{false}, & \text{otherwise} \end{cases} \tag{6}$$

$$I_{B-\text{Half}} = \begin{cases} \text{true}, & \text{if}\left(|S_{i,j} - U_{B-\text{Halfmax}}| \geq Th\_2\right) \\ \text{or } \left(|S_{i,j} - U_{B-\text{Halfmin}}| \geq Th\_2\right) \\ \text{false}, & \text{otherwise} \end{cases} \tag{7}$$

$$\text{Decision 2} = \begin{cases} \text{true}, & \text{if}(I_{T-\text{Half}} = \text{true}) \\ \text{or } (I_{B-\text{Half}} = \text{true}) \\ \text{false}, & \text{otherwise} \end{cases} \tag{8}$$

Perimeter Module. The perimeter module in this algorithm is one such effective mechanism that detects the edges and avoids their filtering unnecessarily. But edge pixels are always of lesser intensity differences computed along the direction of an edge. Since, the only concern is about the centre pixel, arrive at four directions $E1$ to $E4$ passing through the centre pixel using Eqs. (9–13) determine whether the centre pixel is an edge pixel or not. The directional differences are computed as follows:

$$F_{E1} = \begin{cases} \text{false}, & \text{if}\left(|a - S_{i,j}| \geq \text{Th\_3}\right) \\ \text{or } \left(|h - S_{i,j}| \geq \text{Th\_3}\right) \\ \text{or } (|a - h| \geq \text{Th\_4}) \\ \text{true}, & \text{otherwise} \end{cases} \tag{9}$$

$$F_{E2} = \begin{cases} \text{false}, & \text{if}\left(|c - S_{i,j}| \geq \text{Th\_3}\right) \\ \text{or } \left(|f - S_{i,j}| \geq \text{Th\_3}\right) \\ \text{or } (|c - f| \geq \text{Th\_4}) \\ \text{true}, & \text{otherwise} \end{cases} \tag{10}$$

$$F_{E3} = \begin{cases} \text{false}, \quad \text{if}\left(\left|b - S_{i,j}\right| \geq \text{Th\_3}\right) \\ \text{or } \left(\left|g - S_{i,j}\right| \geq \text{Th\_3}\right) \\ \text{or } (|b - g| \geq \text{Th\_4}) \\ \text{true}, \ \text{otherwise} \end{cases} \tag{11}$$

$$F_{E4} = \begin{cases} \text{false}, \quad \text{if}\left(\left|d - S_{i,j}\right| \geq \text{Th\_3}\right) \\ \text{or } \left(\left|e - S_{i,j}\right| \geq \text{Th\_3}\right) \\ \text{or } (|d - e| \geq \text{Th\_4}) \\ \text{true}, \ \text{otherwise} \end{cases} \tag{12}$$

$$\text{Decision 3} = \begin{cases} \text{false}, \quad \text{if}(F_{E1}) \text{ or } (F_{E2}) \\ \text{or } (F_{E3}) \text{ or } (F_{E4}) \text{ is true} \\ \text{true}, \ \text{otherwise} \end{cases} \tag{13}$$

As per decision 3, if any of the directional differences F_{E1} to F_{E4} is true, i.e. shows the existence of an edge, then, Decision 3 is made false, i.e. the centre pixel is uncorrupted, else, the centre pixel is considered as corrupted. Since a parallel logic is used, each module is expected to be independent and self-sufficient, so as to deliver the judgment onto the centre pixel.

Analogous Module. In the analogous module, to identify the noisy pixels first consider two sets of thresholds, out of which only those thresholds that are closer to the general pixel range are chosen. This improves the accuracy of the detection mechanism. To determine the thresholds, first, the nine pixels within 3 × 3 mask are sorted in ascending order. In ascending order, the fifth value obtained is the median represented as $K_{i,j}$, the value preceding it is the fourth value represented as $J_{i,j}$ and the value succeeding it is the sixth value represented as $L_{i,j}$. Now use these values to obtain the thresholds as

$$\begin{aligned} W_{\max} &= L_{i,j} + \text{Th\_5} \\ W_{\min} &= J_{i,j} - \text{Th\_5} \end{aligned} \tag{14}$$

These are the first set of thresholds obtained by adding and subtracting a fixed threshold from the sixth and fourth pixels, respectively. Now, the following set of equations give the final set of thresholds

$$T_{\max} = \begin{cases} W_{\max} \quad \text{if}\left(W_{\max} \leq K_{i,j} + \text{Th\_6}\right) \\ K_{i,j} + \text{Th\_6}, \quad \text{otherwise} \end{cases} \tag{15}$$

$$T_{\min} = \begin{cases} W_{\min} \quad \text{if}(W_{\min} \geq K_{i,j} - \text{Th\_6}) \\ K_{i,j} - \text{Th\_6}, \quad \text{otherwise} \end{cases} \tag{16}$$

Here, $T_{\max}$ and $T_{\min}$ are the final thresholds that are used to test the centre pixel, while $K_{i,j}$ + Th_6 and $K_{i,j}$ − Th_6 are the second set of thresholds. Only those thresholds that are closer to the general pixel range are chosen here. Now, the final decision, whether the centre pixel is noisy or uncorrupted, is arrived at by comparing the centre pixel value with the two final thresholds $T_{\max}$ and $T_{\min}$. If the centre pixel value lies in between the thresholds $T_{\max}$ and $T_{\min}$, then it is considered as uncorrupted; else, it is considered as corrupted and is filtered. This decision can be expressed as

$$\text{Decision 1} = \begin{cases} \text{true}, & \text{if}\left(S_{i,j} \geq T_{\max}\right) \text{ or } \left(S_{i,j} \leq T_{\min}\right) \\ \text{false}, & \text{otherwise} \end{cases} \tag{17}$$

2.2 Edge Preserving Filter

The basic logic behind the working of the edge preserving filter is replacing the centre pixel with a mean value of only those pixels that are involved in the edge. To detect the presence of edge, consider eight directions within the 3 × 3 mask and their corresponding directional differences are D_1 to D_8. The direction with the least directional difference is more likely to have an edge. Thus, the pixels that make up that direction are used to find the mean. To avoid wrong calculation of directional difference, omit all those directions from D_1 to D_8 that include a noisy pixel during the calculation of directional differences. To determine which pixels are noisy, the $W_{\max}$ and $W_{\min}$ from the analogous module are used. In the case where all the neighbours of the centre pixel are suspected to be noisy, the weighted average of a, b and c is to be found. The estimated grey scale value of the centre pixel is given by

$$\hat{S}_{i,j} = (a + b \times 2 + c)/4 \tag{18}$$

Here, $\hat{S}_{i,j}$ denotes the estimated value of the centre pixel. While, 'a, b and c' denote the upper row neighbours of the centre pixel that are filtered previously. Generally, if none of the neighbours were found to be noisy, calculate the edge distances using the equations below

$$\begin{aligned} D_1 &= |d-h| + |a-e|,\\ D_2 &= |a-g| + |b-h|,\\ D_3 &= |b-g| \times 2,\\ D_4 &= |b-f| + |c-g|,\\ D_5 &= |c-d| + |e-f|,\\ D_6 &= |d-e| \times 2,\\ D_7 &= |a-h| \times 2,\\ D_8 &= |c-f| \times 2, \end{aligned} \tag{19}$$

Based on which of the directional differences ends up being the $D_{\min}$, the estimated grey scale value of the centre pixel is calculated as given below

$$\hat{S}_{i,j} = \begin{cases} (a+d+e+h)/4, & \text{if } D_{\min} = D_1\\ (a+b+g+h)/4, & \text{if } D_{\min} = D_2\\ (b+g)/4, & \text{if } D_{\min} = D_3\\ (b+c+f+g)/4, & \text{if } D_{\min} = D_4\\ (c+d+e+f)/4, & \text{if } D_{\min} = D_5\\ (d+e)/4, & \text{if } D_{\min} = D_6\\ (a+h)/4, & \text{if } D_{\min} = D_7\\ (c+f)/4, & \text{if } D_{\min} = D_8 \end{cases} \tag{20}$$

It is to be noted that whenever the median of the pixels b, d, e and g is computed, it always results in a value equal to that of the estimated value. In the case of a wrong detection of an edge or any other causes of error, the estimated value is wrongly calculated. To prevent this, the median of the pixels b, d, e and g along with estimated value is computed and use the resultant median to replace the centre pixel value. The final median value obtained, which replaces the centre pixel, is given by

$$\hat{S}_{i,j} = \text{Median}\left(\hat{S}_{i,j}, b, d, e, g\right) \tag{21}$$

3 Results and Discussion

Above proposed algorithm has been simulated on the MRI images of brain. The proposed algorithm has been implemented using MATLAB R2013a. Experiments are conducted on three images of size 256 × 256 namely Cavernos_Angomia, Cerebral_Hemorrhage and Glioma brain images with 8-bit pixel amplitude resolution. These images are captured from Siemens-Area MRI scanner equipment, in which image is captured on slice of thickness 1 mm × 1 mm × 1 mm by 48 multi channels with 1.5 Tesla magnetic field intensity. The advanced technology used in this equip-

Table 1 Shows comparison of restored Cavernos_Angomia brain image using PSNR for various noise densities

Method	PSNR (dB) values for Cavernos_Angomia			
	5%	10%	15%	20%
Noisy	23.47	17.12	15.56	14.16
Median	33.19	32.53	31.81	30.95
AMF	37.50	33.96	33.47	31.70
Proposed	41.89	39.05	36.95	35.32

Table 2 Shows comparison of restored Cerebral_Hemorrhage brain image using PSNR for various noise densities

Method	PSNR (dB) values for Cerebral_Hemorrhage			
	5%	10%	15%	20%
Noisy	23.33	19.37	17.35	15.07
Median	30.25	29.76	29.07	28.48
AMF	33.19	32.55	30.66	29.30
Proposed	37.45	35.32	33.76	32.27

Table 3 Shows comparison of restored Glioma brain image using PSNR for various noise densities

Method	PSNR (dB) values for Glioma			
	5%	10%	15%	20%
Noisy	21.42	19.67	17.11	15.21
Median	30.74	30.30	29.85	29.30
AMF	33.66	32.59	31.28	30.22
Proposed	37.84	36.01	34.43	33.04

ment is Magnetom Avanto-Tim technology. While simulating, random noise will be added to image to distort it. The density of noise applied on the image vary from 5 to 20%. The performance of above algorithm will be represented in terms of PSNR is given as

$$\mathrm{PSNR} = 10 \log \frac{255^2}{\mathrm{MSE}} \tag{22}$$

where MSE stands for mean square error. The obtained values of PSNR for a Cavernos_Angomia, Cerebral_Hemorrhage and Glioma brain images for various noise densities from 5 to 20% in the steps of 5% have been tabulated and are shown in Tables 1, 2 and 3 respectively.

From results, it can be stated that the proposed algorithm exhibits better PSNR than existing algorithms namely Median and Adaptive Median Filter (AMF). Figures 1, 2 and 3 show bar charts of different methods on various brain images in case of PSNR with respect to various noise levels. From the figures, it is observed that the proposed method has better performance than other methods in all noise levels.

Figures 4, 5 and 6 illustrate the performance of the proposed denoising algorithm for brain images Cavernos_Angomia, Cerebral_Hemorrhage and Glioma at noise density 20, 15 and 10% of random noise respectively to explore the visual quality.

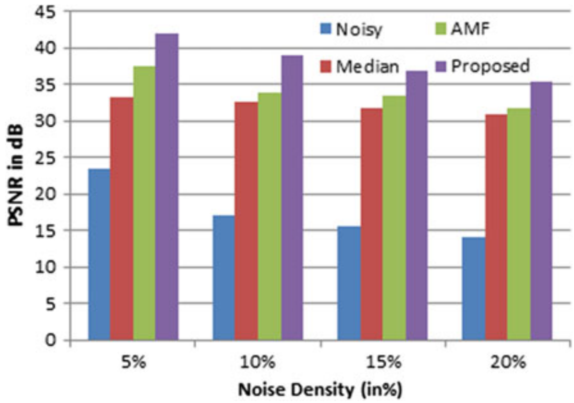


Fig. 1 Evaluation of the PSNR values for various algorithms applied on Cavernos_Angomia brain image for varying noise levels from 5 to 20% of noise

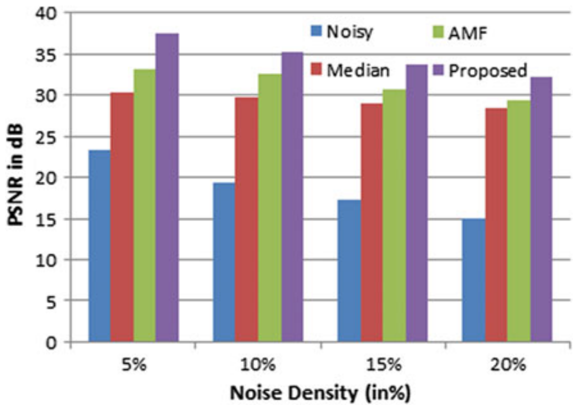


Fig. 2 Evaluation of the PSNR values for various algorithms applied on Cerebral_Hemorrhage brain image for varying noise levels from 5 to 20% of noise

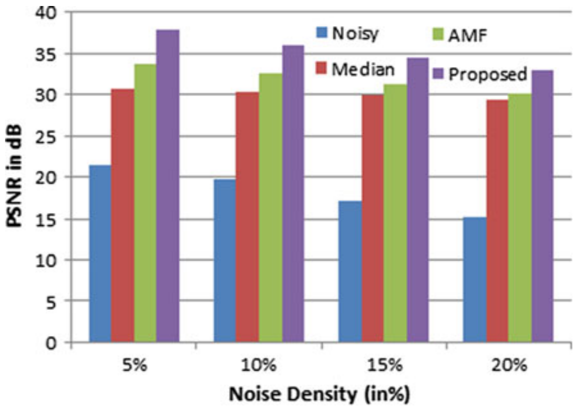


Fig. 3 Evaluation of the PSNR values for various algorithms applied on Glioma brain image for varying noise levels from 5 to 20% of noise

Figures 4, 5 and 6 (i) represent original brain image, Figs. 4, 5 and 6 (ii) represent the noisy image and Figs. 4, 5 and 6 (iii) represent reconstructed image of pro-

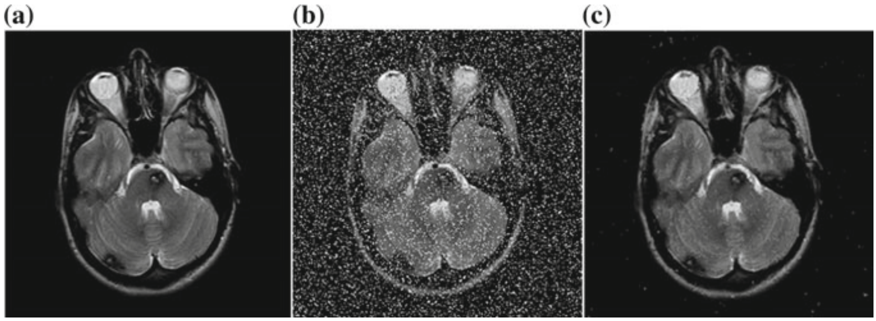


Fig. 4 Performance of the proposed algorithm for Cavernos_Angomia Brain image at a noise density of 20% Random noise **a** original image [13] **b** noisy image **c** denoised image

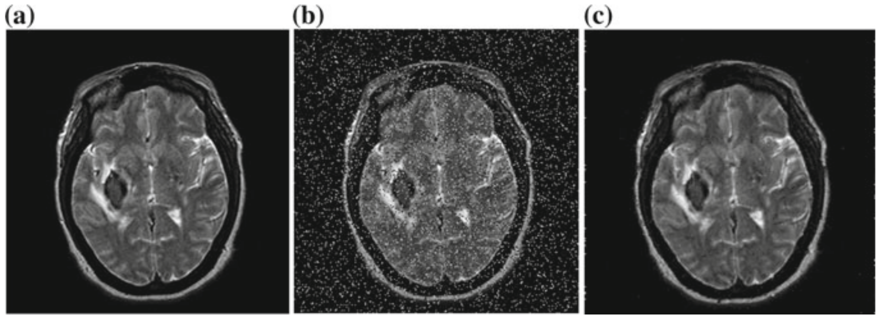


Fig. 5 Performance of the proposed algorithm for Cerebral_Hemorrhage brain image at a noise density of 15% Random noise **a** original image [13] **b** noisy image **c** denoised image

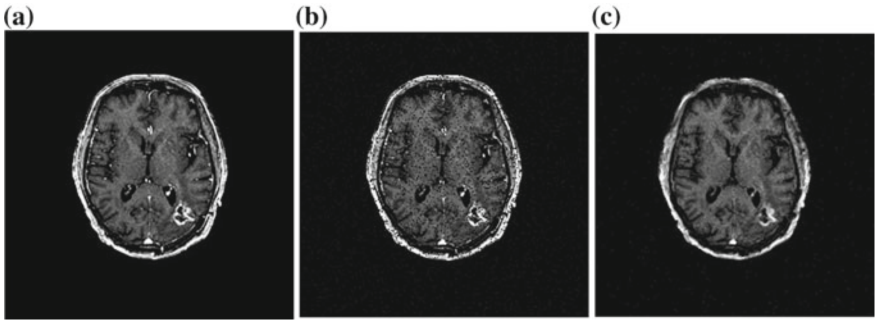


Fig. 6 Performance of the proposed algorithm for Glioma brain image at a noise density of 10% Random noise **a** original image [13] **b** noisy image **c** denoised image

posed denoising algorithm. From Figs. 4, 5 and 6 (iii), it is observed that reformed images obtained by applying the proposed algorithm preserve the edges in process of denoising random-valued impulse noise in an image.

4 Conclusions

In this paper, an decessive algorithm to denoise randomly valued impulse noise from medical images is proposed. The algorithm has been tested on various MRI brain images and produces better and accurate results. The proposed method uses an impulse detector and an edge preserving filter to denoise the corrupted pixels. The method uses simple mathematical expressions to detect and denoise impulse noise in images. Due to this simple approach the proposed method falls under the category of low complexity techniques and provides a better performance in terms of PSNR than that of the low complexity techniques, making it very helpful in telemedicine applications.

References

1. Dong Y, Xu S (2007) A new directional weighted median filter for removal of random valued noise. IEEE Sig Process Lett 14(3):193–196
2. Chen P-Y, Lien C-Y (2008) An efficient edge-preserving algorithm for removal of salt and pepper noise. IEEE Sig Process Lett 15:833–836
3. Yu H, Zhao L, Wang H (2008) An efficient procedure for removing random valued impulse noise in images. IEEE Sig Process Lett 15:922–925
4. Luo W (2006) An efficient detail-preserving approach for removing impulse noise in images. IEEE Sig Process Lett 13(7):413–416
5. Ng P-E, Ma K-K (2006) A switching median filter with boundary discriminative noise detection for extremely corrupted images. IEEE Trans Image Process 15(6):1506–1516
6. Dash A, Sathua SK (2015) High density noise removal by using cascading algorithms. In: Fifth international conference on advanced computing & communication technologies
7. Chan RH, Ho CW, Nikolova M (2005) Salt-and-Pepper noise removal by median-type noise detectors and detail-preserving regularization. IEEE Trans Image Process 14(10):1479–1485
8. Sun T, Neuvo Y (1994) Detail-preserving median based filters in image processing. Pattern Recogn Lett 15:341–347
9. Santhanam T, Chithra K (2014) A new decision based unsymmetric trimmed median filter using Eucledian distance for removal of high density salt and pepper noise from images. In: IEEE conference publications, pp 1–5
10. Suman S (2014) Image denoising using new adaptive based median filter. Int J (SIPIJ) 5(4):1–13
11. Vikrant B, Tiwari H, Srivastava A (2015) A non-local means filtering algorithm for restoration of Rician distributed MRI. In: Emerging ICT for bridging the future—proceedings of the 49th annual convention of the Computer Society of India CSI, vol 2. Springer, Cham
12. Awanish KS, Vikrant B, Verma RL, Alam MS (2014) An improved directional weighted median filter for restoration of images corrupted with high density impulse noise. In: 2014 international conference on optimization, reliability, and information technology (ICROIT). IEEE
13. https://www.frontiersin.org/articles/10.3389/fneur.2015.00033/full

Importance of LEDs Placing and Uniformity: Phototherapy Treatment Used for Neonatal Hyperbilirubinemia

J. Lokesh, K. Shashank and Savitha G. Kini

Abstract Phototherapy devices are most popularly used in medical applications to treat jaundice in infants called Neonatal jaundice. Neonatal jaundice or neonatal hyperbilirubinemia is the symptom of excessive bilirubin in blood in 60% of term babies and almost up to 70% of preterm babies. In earlier studies, it is found that the efficacy of phototherapy devices depends on spectral wavelength, body surface area, irradiance and duration of exposure. And it is important to know even importance of uniformity, peak wavelength and optimum height contribution in reducing phototherapy duration time. The paper explains the importance of uniformity and the optimum height. The uniformities for both the sets of LEDs are determined and selected for further testing. The optimum height at which the LEDs must be mounted is determined by comparing the uniformity values. The temperature is compared with the standard maximum ambient temperature up to which the infant remains healthy.

1 Introduction

Jaundice is a condition in which blood contains an excess amount of unconjugated bilirubin. Almost 60% term babies and 70% preterm babies have this condition. It is usually detected within 24–48 h of birth. Phototherapy is the widely used treatment. Earlier people would just expose the infant to sunlight for a certain period of time. But it was later found out that only light in the wavelength range of 430–490 nm was effective for the conversion of unconjugated bilirubin. In particular, 450–460 nm was found to be most effective. Most of the phototherapy devices use either FTLs or

J. Lokesh (✉) · K. Shashank · S. G. Kini
Electrical and Electronics Department, Manipal Institute of Technology, MAHE, Manipal 576104, India
e-mail: lokesh.j@manipal.edu

K. Shashank
e-mail: shanky.rao7@gmail.com

S. G. Kini
e-mail: savitha.kini@manipal.edu

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_38

blue CFLs or a combination of both. The disadvantages of FTL devices are that the light sources give light at broader range of frequencies, thus reducing efficacy. CFLs and FTLs dissipate a large amount of heat because of ballasts. This also makes them very bulky, thus hindering mobility. Also, the lifespan of these lamps is very less, about 1000–1500 h. Apart from these, there are also health issues upon continuous exposure of infants to the mentioned light sources. The presence of mercury also raises concern because any exposure is very harmful for the infant. And the risk of exposure is very high in poorly maintained lamps.

Frequent exposure of infants to FTLs may cause Irlen Syndrome, a perception problem which affects the infant's ability to read and write in the future. In very rare cases, exposure to FTLs is also known to induce migraines. Stony Brook University researchers conducted a survey of CFLs bought from in and around their locality. The integrity of the phosphor coatings of each lamp was tested. They observed cracks in all the bulbs, through which UVA and UVC radiation were leaking out.

At Stony Brook's Advanced Energy Research and Technology Center (AERTC), the effects of these same bulbs on human skin tissue were tested. Comparisons were made with incandescent bulbs of the same intensity. It was found out that the effect of UV radiations emitted from the CFLs on human skin tissue was consistent with damage from exposure to ultraviolet radiation while there was no adverse effect of incandescent bulbs on human skin. These problems urge the need to use other sources of light for phototherapy as the effect might be magnified on an infant. But recently, LEDs are slowly replacing them because of low size and longer life. There is also no UV radiation emission and no mercury in LEDs. The major advantage of LED is that the exact wavelength required for a specific purpose can be obtained, with very small deviations in the wavelength. This paper provides the design of an LED-based phototherapy device.

2 Literature Review

Jaundice is a commonly encountered problem after birth and occurs in 60–70% of term infants and nearly all of preterm infants, including those near-term infants 35–38 weeks gestational age [1]. Nomogram for designation of risk in infants based on bilirubin concentration in blood at different ages is shown in Fig. 1.

For treatment of hyperbilirubinemia, light, in the range of approximately 400–500 nm with a peak at 460 $\pm$ 10 nm, is considered the most effective. In the current AAP guideline, intensive phototherapy is defined as the use of blue light (in the 430–490 nm band) delivered at 30 W/cm^2/nm or higher to the greatest BSA as possible. The light source must be placed at a height of 40 cm from the infant for optimal result. An irradiance level of 8–10 μW cm^{-2} nm^{-1} is considered to be normal phototherapy. Irradiance of >30 μW cm^{-2} nm^{-1} is considered to be Intensive phototherapy [1–4]. The Phototherapy enables the body to get rid of excess bilirubin through stools. It is also seen that while irradiance in the range of 30–40 μW cm^{-2}

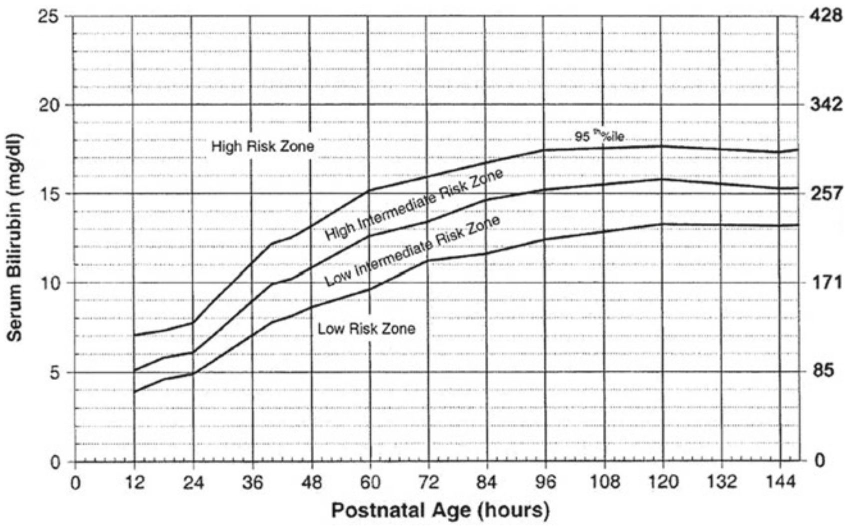


Fig. 1 Nomogram for designation of risk in infants based on bilirubin concentration in blood at different ages (*Source* AAP Guideline [1])

nm^{-1} is optimal, the rate of degradation of bilirubin is linear up to an irradiance of 55 $\mu W\ cm^{-2}\ nm^{-1}$ with no saturation point [2].

The normal acceptable range of Total Serum Bilirubin (TSB) in an infant is 5.2 mg/dL [5]. Phototherapy is administered to an infant if the TSB levels are more than this value. The type of phototherapy varies between normal and intensive phototherapy. If the TSB value is <15 mg/dL, then normal phototherapy with an irradiance of 8–10 $\mu W/cm^2/nm$ is administered and intensive phototherapy with an irradiance of >30 $\mu W/cm^2/nm$ is administered if the TSB level is >20 mg/dL. For TSB values >25 mg/dL, blood exchange transfusion is recommended [6].

It is observed that efficiency of phototherapy is dependent on the wavelength, the comparative study between BSL (Broad spectrum light) (420–680 nm, and peak wavelength of 455 and 524 nm) and Blue LED (400–500 nm and peak of 460 nm) shows the duration of phototherapy was lower in the BSL than in the blue LED phototherapy group (15.8, 4.9 vs. 20.6, 6.0 h; $p = 0.009$) [4, 7, 8]. Blue plus green phototherapy is as effective as blue phototherapy and it attenuates irradiation-induced oxidative stress. And it is observed that the oxidative stress induced by irradiation level can be attenuated by adding green spectrum with blue and is as effective as only blue spectrum to reduce bilirubin level and tested in gunn rat model [9–13]. From the above discussion, it is seen that how the efficacy of phototherapy is dependent on wavelengths and irradiation level. In this paper, the importance of uniformity, peak wavelength and optimum height contribution in reducing phototherapy duration time and improving of efficiency of phototheapry is discussed.

3 Methodology and Validation

The prototype device is of glass of dimensions 56 cm × 49 cm × 27 cm. The inner walls and the base of the prototype are lined with light blue cardboard sheet to obtain highest possible reflectance. The base of the prototype is marked with a rectangle of 40 cm × 30 cm which is divided into 36 grid points, as shown in Fig. 2.

The LEDs are placed along the sides of the prototype and controlled with an 8 channel Optronix driver module and GUI window is used to set the input current required for the operation for good uniformity and right peak wavelength. The current, wavelength, irradiance level is obtained by testing LEDs in spectrometer (Integrating sphere). Two different set of LEDs are selected and position height is determined for better uniformity, as height changes, irradiance level and uniformity also changes. The current supplied to the LED of set 1, which are high brightness (HB) LEDs, is varied from 15 to 350 mA in steps of 10 mA and to the LED of set 2, which are low brightness (LB) LEDs, is varied from 100 to 350 mA in steps of 5 mA is shown in Tables 1 and 2, respectively, for single LED.

The selected irradiance level for test purposes is 55 μW cm^{-2} nm^{-1}. It is so selected because this is the maximum irradiance value which can be used without any change in the rate of degradation of bilirubin [1]. From Eqs. (1) and (2), we determine the required luminous flux of each LED as 7.65 Lm. From Tables 1 and 2, the current required to obtain this value of luminous flux from each HB LED is 75 and 318 mA for each LB LED.

$$\Phi_{\text{total}} = A * E_{e,\lambda} \tag{1}$$

$$\Phi = \Phi_{\text{total}}/N \tag{2}$$

where

Φ_{total} is the total radiant power;
A is the illuminated area;

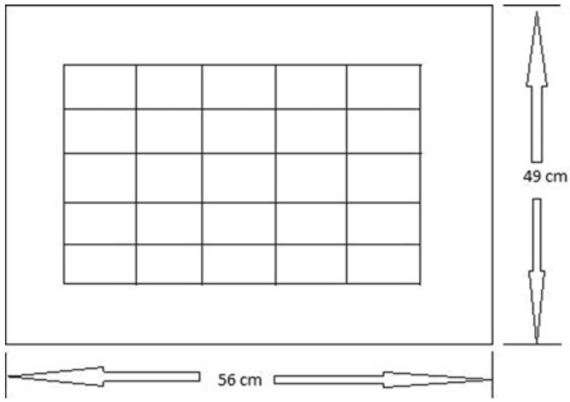


Fig. 2 Depiction of grid points on prototype base

Table 1 HB LED: forward current, wavelength, and luminous flux

I_f (mA)	λ (nm)	Φ (Lm)
15	451	1.324
45	451	3.825
75	**450**	**7.663**
105	450	8.727
135	450	10.18
165	450	12.44
195	450	14.22
225	450	17.11
255	450	19.08
285	450	21.04
315	450	22.92
345	450	25.40
350	450	25.76

Table 2 LB LED: forward current, wavelength, and luminous flux

I_f (mA)	λ (nm)	Φ (Lm)
100	453	3.481
125	453	4.178
150	452	4.483
175	452	4.937
200	451	5.492
225	451	5.791
250	451	6.007
275	451	6.606
300	450	7.324
318	**450**	**7.689**
325	450	7.749
350	450	8.243

$E_{e,\lambda}$ is the irradiance level selected;
Φ is the radiant power of each LED; and
N is the number of LEDs.

Prototype: Along the length of the prototype, the LEDs are placed 20 cm from either corner. Along the width of the prototype, LEDs are placed 16 cm from either corner. They are placed at a height of 22 from the workplace. The illuminance values at the grid points on the workplace are noted. This process is done for the LED placement heights of 23 and 24 cm also and the comparison of uniformity of 2 sets of LEDs for different heights is given in Table 3 with the same number (8 Nos) of LEDs.

The uniformity is calculated from Eq. (3).

$$U = E_{min}/E_{avg} \quad (3)$$

where

E_{min} is minimum illuminance
E_{avg} is average illuminance
U is uniformity

From Table 3, even though the LB LEDs offer higher uniformity, the current required is much higher than the current required for the HB LEDs, which is not desirable and HB LEDs are the most suitable for use because the difference between the uniformities is not large at a given height. And it is also observed that variations in numbers (Number of LEDs) gives different uniformities. The uniformity level with multiple (8, 6 and 4 nos) HB LEDs is shown in Table 4 and the illuminance distribution with 8 and 6 nos of LEDs are shown in Figs. 3 and 4, respectively. And illuminance distribution with 4 LEDs is not shown as it gives low uniformity.

The uniformity is highest at a height of 23 cm from the workplace with very little variation in the illuminance and the illuminance values are higher at the center region

Table 3 Height from the workplace, uniformity of HB LEDs, uniformity of LB LEDs

Height (cm)	U_{HBLED}	U_{LBLED}
22	0.7582	0.8588
23	0.809	0.8892
24	0.7822	0.8674

Table 4 Height from workplace, uniformities with 8, 6, and 4 LEDs

Height (cm)	U_{8LED}	U_{6LED}	U_{4LED}
22	0.7582	0.73	0.659
23	**0.809**	**0.757**	**0.664**
24	0.7822	0.721	0.631

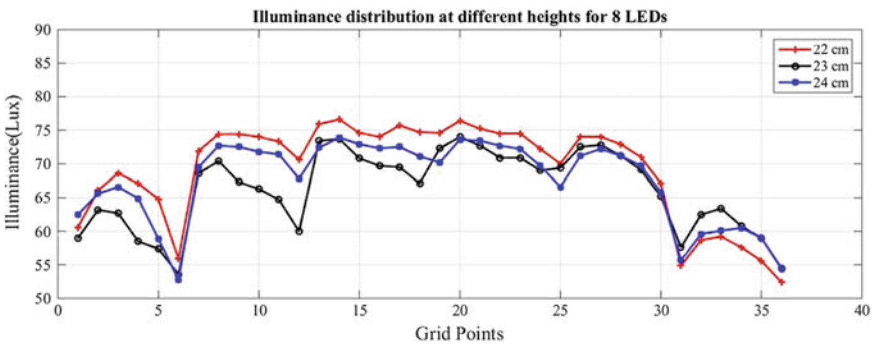


Fig. 3 Illuminance distribution at different heights for 8 HB LEDs

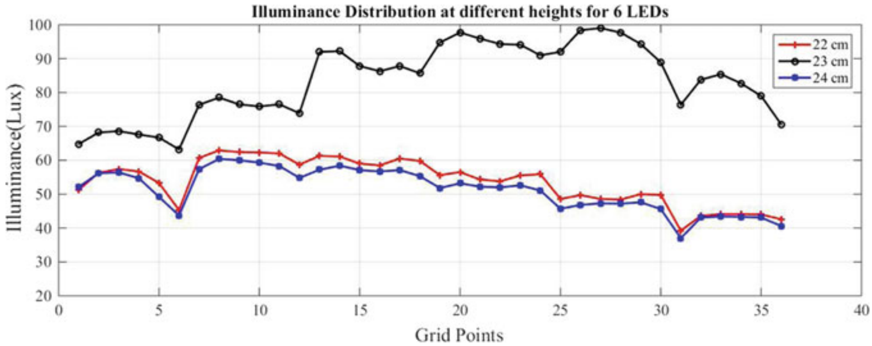


Fig. 4 Illuminance distribution at different heights for 6 HB LEDs

Table 5 Uniformity versus phototherapy: total duration (h)

	Irradiance (μW cm^{-2} nm^{-1})	Uniformity	Phototherapy: total duration (h)
Available	35	0.62	11.48
Proposed	35	0.809	11.39

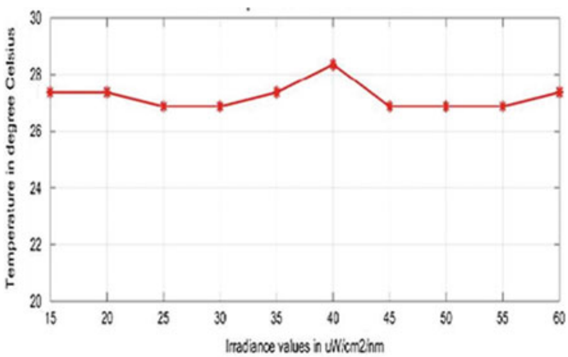


Fig. 5 Irradiance values and maximum temperature at 23 cm height

of the workplace for all the three heights (22, 23 and 24 cm), since at sufficiently greater heights, the light spread is more even at the center of the workplace and the main area of the workplace is adequately illuminated, avoiding any dark.

And case study shows uniformity will also plays an important role in reducing bilirubin level at bit less duration is shown in Table 5.

As efficacy of phototherapy is dependent on the irradiance level and irradiance level exposed is dependent on initial bilirubin, it is important to check the temperature variations for a change in irradiance level to avoid forced phototherapy and skin problems due to more irradiance. The results are shown in Fig. 5 and it is observed that there is no much changes in temperature for change in irradiance level in the proposed system as LEDs are placed at different points to maintain better uniformity. The accepted ambient temperature for an infant to not be adversely affected is 32 °C. The maximum temperature obtained at different irradiance values varies in a small interval of 26.88–28.35 °C, which is within the ambient temperature mentioned. This temperature will be even lower in an actual hospital where the room for the infants is temperature controlled with air-conditioning.

4 Conclusion

The illuminance distribution on the workplace with different heights of LED placement was found out and the most suitable position to mount LEDs was determined based on the uniformity values. Furthermore, comparisons of uniformity were done for 2 different sets of LEDs. And it is shown that importance of uniformity in reducing the treatment duration to avoid skin problems and other. The temperature at the workplace was also tested and the maximum temperature for different values of irradiance is noted, which varies in a very small interval and is less than the recommended ambient temperature. Future study is to know will peak wavelength also have an impact in reducing bilirubin level.

References

1. American Academy of Pediatrics (2004) Management of hyperbilirubinemia in the newborn infant 35 or more weeks of gestation. Pediatrics 114:297–316
2. Vreman HJ, Wong RJ, Stevenson DK (2004) Phototherapy: current methods and future directions. Semin Perinatol 28:326–333 (Elsevier Inc.)
3. Ennever JF, McDonagh AF, Speck WT (1983) Phototherapy for neonatal jaundice: optimal wavelength of light. J Pediatr 103:295–299
4. Seidman DS, Moise J, Ergaz Z et al (2003) A prospective randomized controlled study of phototherapy using blue and blue-green light-emitting devices, and conventional halogen-quartz phototherapy. J Perinatol 23:123–127
5. de Araujo MCK, Vaz FAC, Ramos JLA (1996) Progress in phototherapy. Sao Paulo Med J 114(2):1134–1140
6. Watchko JF, Maisels MJ (2003) Jaundice in low birthweight infants: pathobiology and outcome, pp F455–459
7. Pratesi S, Di Fabio S, Bresci C, Di Natale C, Bar S, Dani C (2014) Broad-spectrum light versus blue light for phototherapy in neonatal hyperbilirubinemia: a randomized controlled trial. Perinatology. ISSN 0735-1631
8. Ebbesen F, Madsen P, Støvring S, Hundborg H, Agati G (2007) Therapeutic effect of turquoise versus blue light with equal irradiance in preterm infants with jaundice. Acta Paediatr 96(6):837–841
9. Uchida Y, Morimoto Y, Uchiike T, Kamamoto T, Hayashi T, Arai I, Nishikubo T, Takahashi Y (2015) Phototherapy with blue and green mixed-light is as effective against unconjugated jaundice as blue light and reduces oxidative stress in the Gunn rat model. Elsevier Early Hum Dev 91:381–385

10. Vecchi C, Donzelli GP, Migliorini MG, Sbrana G (1983) Green light in phototherapy. Pediatr Res 17(6):461–463
11. Mohammadizadeh M, Eliadarani FK, Badiei Z (2012) Is the light-emitting diode a better light source than fluorescent tube for phototherapy of neonatal jaundice in preterm infants? Adv Biomed Res 1:51
12. Uchida Y, Morimoto Y, Haku J, Nakagawa T, Nishikubo T, Takahashi Y (2011) A comparison the therapeutic effects of blue light-emitting diodes (LED) and green fluorescence lights (green FL) in the treatment of neonatal jaundice. J Jpn Soc Premature Newborn Med 23(2):263–267
13. Itoh S, Onishi S, Isobe K, Manabe M, Yamakawa T (1987) Wavelength dependence of the geometric and structural photoisomerization of bilirubin bound to human serum albumin. Biol Neonate 51(1):10–17

Automated Glaucoma Detection Using Global Statistical Parameters of Retina Fundus Images

Prathiksha R. Puthren, Ayush Agrawal and Usha Padma

Abstract Glaucoma is an eye disorder which is prevalent in the ageing population and causes irreversible loss of vision. Hence, computer-aided solutions are of interest for screening purposes. Glaucoma is indicated by structural changes in the Optic Disc (OD), loss of nerve fibres and atrophy of the peripapillary region of optic disc in retina. In retina images, most changes appear in form of subtle variation in appearance. Hence, automated assessment of glaucoma from colour fundus images is a challenging problem. Prevalent approaches aim at detecting the primary indicator, namely, the optic cup deformation relative to the disc and use the ratio of the two diameters in the vertical direction, to classify images as normal or glaucomatous. An attempt is made to detect glaucoma by combining image processing and neural network techniques. The risk of blindness can be reduced by 50% with screening patients vulnerable to eye diseases specially glaucoma. The global statistical features of the dataset images are used to detect images as glaucoma or normal. The technique involves screening for the vital signs such as intensity values in the fundus image for detecting glaucoma in patients. The result shows the feasibility of detection of glaucoma for vulnerable patient.

1 Introduction

Glaucoma is a retinal disease and is among the leading causes which results in loss of sight. It is a degenerative optic neuropathy which results in gradual loss of retinal nerve fibre which cannot be revitalized. Thus, untreated glaucoma has a potential to cause irreparable damage to retina. This irreversible and asymptomatic nature

P. R. Puthren · A. Agrawal (✉) · U. Padma (✉)
Department of Telecommunication Engineering, RV College of Engineering, Bengaluru, India
e-mail: ayush.agrawal199508@gmail.com

U. Padma
e-mail: ushapadma2018@gmail.com

P. R. Puthren
e-mail: prathiksha.puthren@gmail.com

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_39

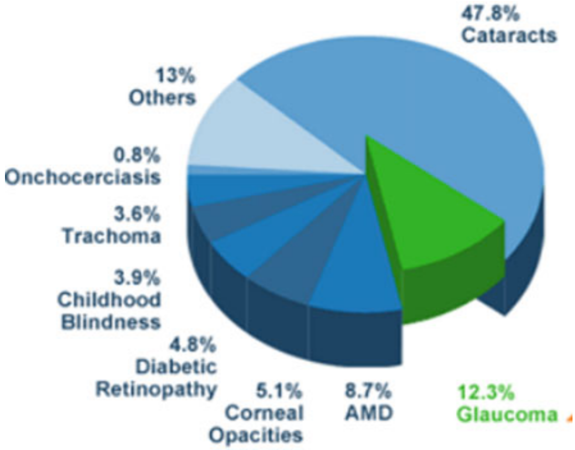


Fig. 1 The chart shows the prevalence of vision impairment in India. *Source* Glaucoma Society of India, p. 55

of the disease emphasizes the need for its timely detection as detecting it in early phases helps in curbing its progression through proper medication. The proliferation of glaucoma cases in recent years adds to the growing concern towards the treatment of the disease. It prevails mainly in ageing population of urban regions. It is estimated to affect 79 million people in the world by the year 2020, showing a 33% increase in the numbers within a decade. Thus, screening for glaucoma is crucial, owing to the nature, for the early detection and enabling effective treatment in early stages to prevent permanent blindness. According to the Glaucoma society of India, glaucoma is the second main cause of blindness in India which is depicted in Fig. 1. The major challenge posed by Indian population is the number. At present, 12 million people in India are affected by glaucoma which is expected to increase to 16 million by 2020. The number of patients per ophthalmologist is around 2–3 lakhs in India. Thus, apart from cost, lack of manpower in terms of skilled technicians poses major challenge in such scenarios. Cost-effective computer-based diagnostic systems can reduce this requirement to a great extent and assist medical experts in diagnosis. Automated screening systems based on retinal (fundus) image analysis can aid in reducing time and effort wasted on analysis of healthy people. Such systems can classify a given case as normal (free of glaucoma-related symptoms) or glaucomatous. Consequently, patients deemed suspect by the system need to be referred to an ophthalmologist. Thus, the categorization of retinal images as normal and glaucomatous is a problem of clinical significance in population screening.

Several techniques for glaucoma detection have been investigated by authors based on available dataset images. In [1], artificial neural network (ANN) techniques were used to detect glaucoma in images. With the help of MATLAB, the disease was classified as normal, severe and mild. Using a neural network, the parameters extracted from MATLAB are compared with standard values. The use of ANN made the detection of Glaucoma accurate as well as adaptive. Ease of operation is the main advantage in this method. In [2], authors implemented algorithm to detect glaucoma which used morphological methods to acquire two main features for identification of Glaucoma, i.e. Ratio of areas of neuro-retinal rim (NRR) in Inferior, Superior, Temporal and Nasal (ISNT) quadrants and Cup to Disc Ratio (CDR). Using this method, the achieved accuracy was 83.5%. In [3], authors reviewed several methods which are useful in identifying glaucoma. By using these techniques, there is need to develop cost-effective automatic techniques for detection of glaucoma disease precisely. In [4], algorithmic rules for robust and automatic extraction of features in colour fundus images were proposed. To locate Optic Disc, Principal Component Analysis (PCA) is used. An active shape model was designed in detection of shape of Optic Disc (OD). Based on fovea localization, fundus coordinate system was made. An approach to identify exudates was by combining edge detection and region growing was suggested. The success percentages of disc boundary detection, disc localization and fovea localization were 94, 99 and 100%, respectively. The specificity of exudates was 71% and sensitivity of the exudates detection was 100%. A screening technique for glaucoma detection by the use of super pixel categorization on optic cup segmentation and Optic Disc was developed [5]. The automated OD segmentation quality was determined by the use of a self-assessment accuracy score. For segmentation of optic cup and for obtaining histograms, location information is included to increase the performance. The results showed an overlapping error of 24.1 and 9.5% in cup and disc segmentation, respectively. In this paper, we have proposed a system for glaucoma detection by combining image processing and neural network techniques. The global statistical parameters of image such as mean, standard deviation and variance are extracted and used for classifying images as normal or glaucoma. The technique used for classifying is the back propagation neural network which achieves high accuracy for classification in several cases.

2 The Proposed Approach

In this section, the dataset, algorithm for detection of ROI and classification of retina fundus images is discussed. Figure 2 shows the generic process of detection of glaucoma.

2.1 Dataset Collection

For developing algorithm for detecting glaucoma, the first necessary step required was obtaining the applicable datasets. 14 fundus images of retina were acquired from different online databases for this purpose. Out of these, 8 images were taken from FAU database and 6 from optic-disc.org database (Fig. 3).

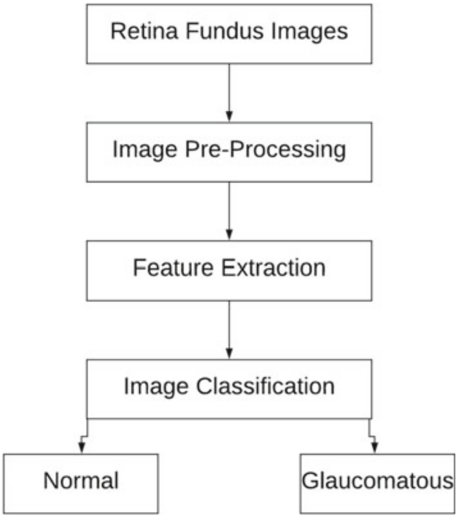


Fig. 2 Generic process of detection of glaucoma

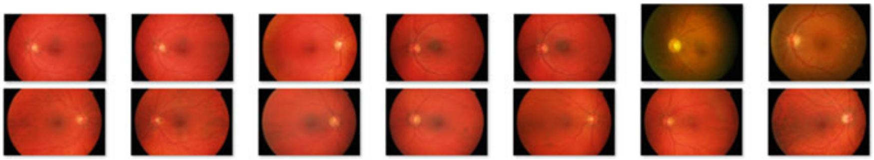


Fig. 3 Dataset images

2.2 *Detection of ROI*

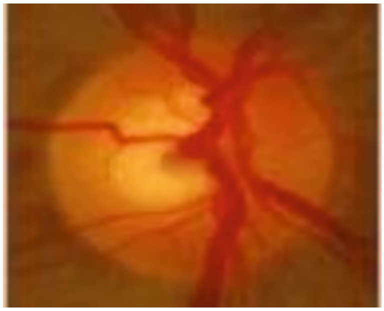
2.2.1 Image Pre-processing

In colour retina fundus images, OD is the most intensified part having orange and pink colour. It is determined as region of interest. The boundary of this Optic Disc must be identified to separate and extract it from the image. The disc occupies around 5% pixels in coloured retina image. The cup and disc extraction was performed on full image and region of interest is located to improve accuracy and reduce cost computation. The region of interest is identified, cropped and resized from all original images as shown in Fig. 4.

2.2.2 Optic Disc and Cup Extraction from Fundus Images

For suspecting glaucoma, assessment of features is an important step, which is determined by optic cup and disc extraction. Initially, the fundus retina image was cropped

Fig. 4 Optic Disc after extraction and resizing



to get OD, and then resized. Then, the blood veins were removed from the image to get a smooth image.

To accomplish this task, morphological operations known as dilation and erosion are performed according to Eqs. (2.1) and (2.2). Dilation is used to increase the size of objects by adding pixels to the borders of object in input image. The structuring element 'DISC' was used for dilating the image. The dilation results in lighting blood vessels and filling internal gaps. Increase in the size of disc also affects CDR. Erosion of image was carried out using same sized structuring element after dilation. Images are eroded for contrasting the boundaries of object. These operations resulted in smooth images without blood vessels.

Dilation is defined as

$$A \oplus B = \bigcup_{b \in B} A_b \tag{2.1}$$

Erosion is defined as

$$A \ominus B = \bigcup_{b \in B} A_{-b} \tag{2.2}$$

where

A: Binary image and B: Structuring element.

After analysing many images, it was concluded Optic Disc had a better contrast in V plane extract from Hue Saturation Value (HSV) images. The mean values of the V plane images were calculated, and given as threshold to convert to binary images. Unwanted objects which were obtained in binary image are removed by labelling them and morphological operations are applied to remove the objects that had fewer pixels, and hence this resulted in removing all the unwanted objects excluding Optic Disc. Then the image boundaries were smoothened by the application of Gaussian filter as depicted in Fig. 5.

As depicted in Fig. 5a, compared to other regions of retina fundus image, a cup has brighter contrast. In the second step, using global threshold which chooses the threshold to decrease the intraclass variance of the black and white pixels, the green

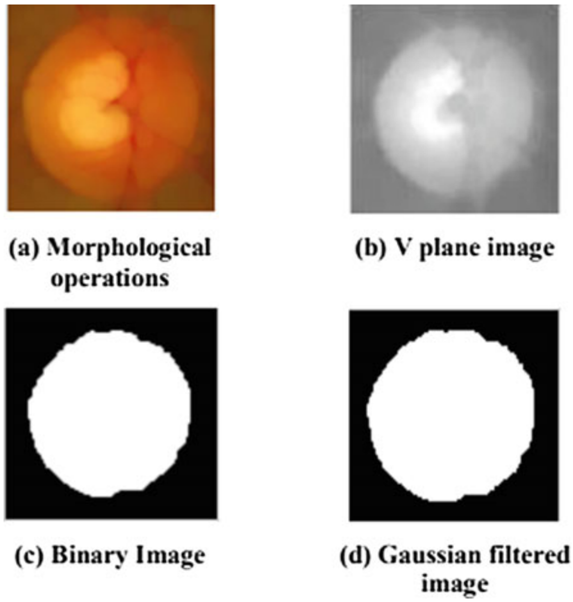


Fig. 5 Extraction of Optic Disc

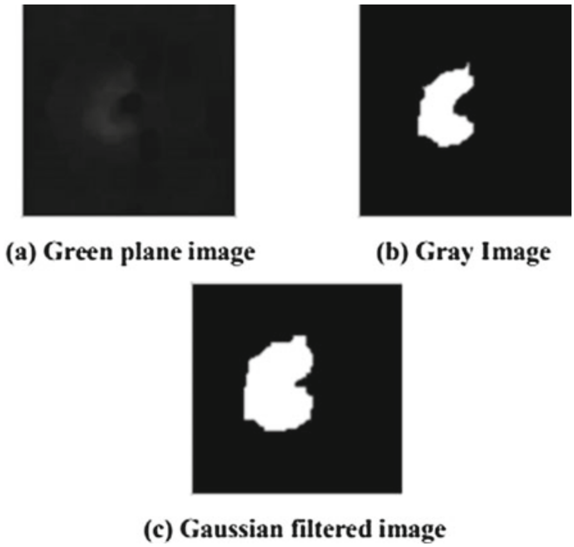


Fig. 6 Optic cup extraction

plane is converted into grey scale image. In the third step, morphological operation for removal of small objects is applied, same as Optic Disc but with less pixel value as cup size is small, this is done after extracting the binary optic cup. Gaussian filter is applied to the resultant binary image of the optic cup, to smoothen the boundaries of optic cup, as shown in Fig. 6.

2.3 Extraction of Features

Basically, feature extraction aims at decreasing the resources needed for describing a large set of necessary data. Thus, the input image can be transformed into set of features. The features extracted from retina fundus images are based on Average Brightness (R, G, B), Standard Deviation (R, G, B) and Coefficient of Variance (R, G, B).

2.3.1 Average Brightness

The sample mean of the pixel brightness within that region is called average brightness. The average brightness (is denoted by m_a) over $\wedge$ pixels inside a region (R) is given as:

$$m_a = \frac{1}{\wedge} \sum_{(m,n) \in R} a[m,n] \tag{2.3}$$

On the other hand, a formulae based on the brightness histogram (unnormalized), $h(a) = \wedge * p(a)$, with discrete brightness values 'a', can be used and is given by

$$m_a = \frac{1}{\wedge} \sum_{a} a \,.\, h[a] \tag{2.4}$$

Here, average brightness (m_a) is an estimate of mean brightness of the brightness probability distribution denoted by $h[a]$.

2.3.2 Standard Deviation

The standard deviation unbiased estimate denoted by s_a, of the brightness's within region (R) for '$\wedge$' pixels is known as sample standard deviation. It is given as:

$$s_a = \sqrt[2]{\frac{\frac{1}{\wedge - 1} \sum_{m,n \in R} (a[m,n] - m_a)^2}{1}} \tag{2.5}$$

Using the histogram formula, the above equation can be written as

$$s_a = \sqrt{\frac{\left(\sum_a a^2 \cdot h[a]\right) - \Lambda \cdot m_a^2}{\Lambda - 1}} \tag{2.6}$$

Here, standard deviation (s_a) is estimate of underlying brightness probability distribution.

2.3.3 Variance

The coefficient of variation (CV) is given by

$$CV = \frac{s_a}{m_a} \times 100\% \tag{2.7}$$

2.4 *Classification*

Classification of the image is carried out based on three features, i.e. mean, standard deviation and variance of the image. Images are sent as input to the neural network which classifies based on the above features. The neural network model used is scaled conjugate gradient back propagation which is discussed in the next section.

2.4.1 Backpropagation Algorithm

Backpropagation is a supervised learning algorithm of ANNs using gradient descent. It is short for 'backward propagating of errors'. If the ANN and error function is given, this method calculates gradient of error function w.r.t neural network's weights. The calculation of the gradient happens backwards in the network. The gradient of last layer of weights is calculated first and the gradient of first layer weights is calculated last. Calculations of one layer are used to calculate gradient of the previous layer.

Backpropagation neural networks are becoming popular in the image and speech recognition applications. It is very efficient algorithm which takes advantage of specialized Graphics Processing Units (GPUs) to improve performance further.

3 Results

A multilayer perceptron backpropagation neural network is used with a hidden layer provided. Backpropagation algorithm based on 9 input neuron and 1 output neuron at the output layer was trained by a 2-layer feed forward network. 28 iterations were required to obtain the best trained network. The trained network with best validation performance had a mean square error (MSE) of 0.000000186 at epoch 28. Result obtained from the confusion matrix showed overall accuracy of 84.6% is achieved using this technique.

In Fig. 7, veins have been extracted by converting the retina fundus image to black and white image. It is mainly done so that, the optical disc which has to be cropped for further image processing does not contain any veins.

In Fig. 8, retina fundus images had been loaded. Average intensity displays the average RGB colour intensity variation. Variance displays the RGB pixels colour

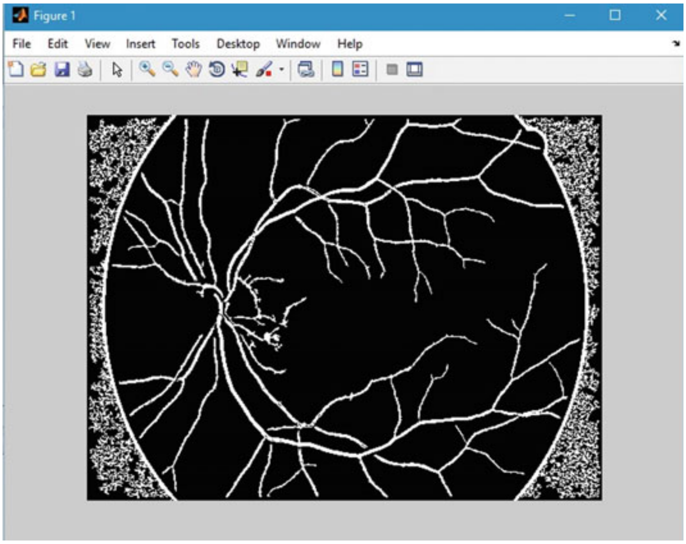


Fig. 7 Extraction of veins

Fig. 8 Optic Disc extraction display window

variance. Median displays the median of RGB pixels intensity. 'Std' displays the standard deviation of RGB colour pixels. Centroid displays the centroid of Optic Disc.

Figure 9 contains the final GUI (graphical user interface) which displays the output whether the retinal fundus is normal or whether it is affected by Glaucoma. It displays mean values, variance values and standard deviation of RGB Colour Pixels.

Fig. 9 Detected Glaucoma for a particular retina fundus image

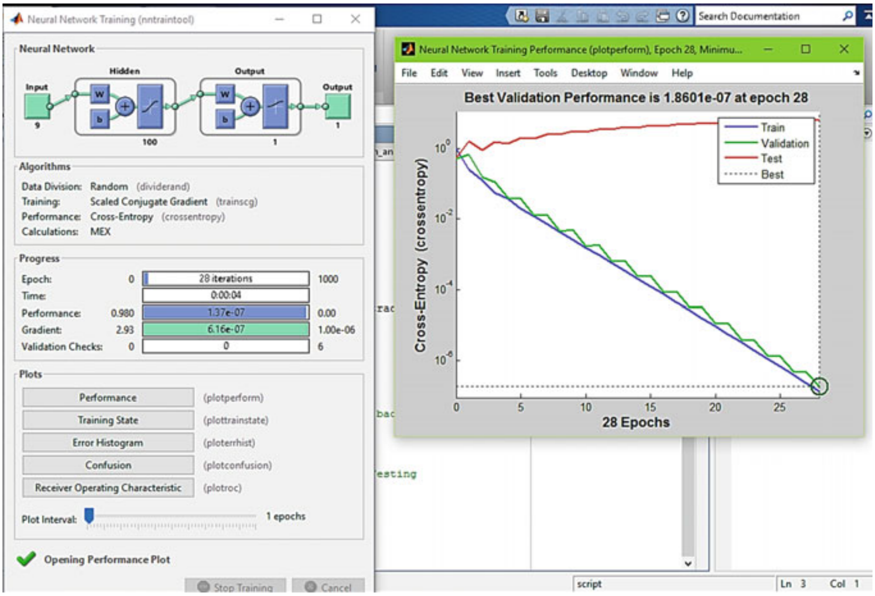


Fig. 10 Display window showing neural network training process with performance plot

The display window in Fig. 10 shows the neural network training model. The performance plot infers that the best validation performance occurs at 28th iteration.

Figure 11 displays the neural network confusion matrix.

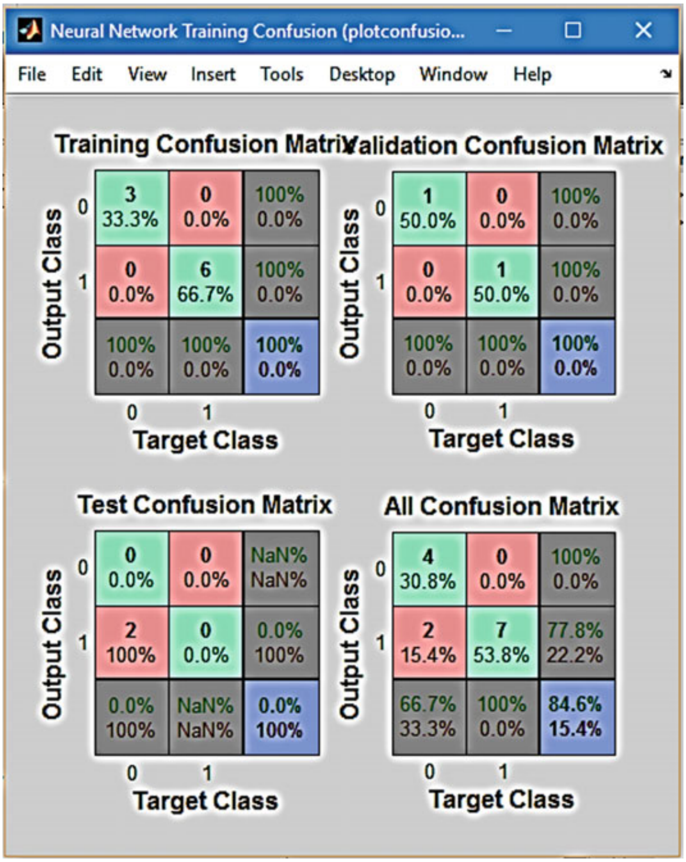


Fig. 11 Neural network training confusion matrix

4 Conclusion

In this paper, an automated glaucoma detection technique with three features, i.e. mean intensity, standard deviation and variance of the dataset images, which facilitates diagnosis of glaucoma is discussed. Morphological techniques were used to extract the features in this novel method. To observe the performance of this approach, 14 retina fundus images were used and processed. The method achieved an accuracy of around 84.6% using global statistical features of dataset images. The results are subjective, fast and consistent even though the performance is far from perfect. The results do not rely on trained glaucoma specialists or the use of specialized, costly OCT/HRT machines. The better performance of this method results in a large scale clinical evaluation and will be helpful in reporting large clinical findings in the future.

Acknowledgements This work was supported by Department of Telecommunication, R.V. College of Engineering, Bangalore, India. The authors would like to thank the Department for providing excellent facilities and timely guidance throughout the completion of the project.

References

1. Naveen Kumar B, Chauhan RP, Dahiya N (2016) Detection of Glaucoma using Image processing techniques: a review. IEEE Trans Biomed Eng 62(5)
2. Sheeba O, George J, Rajin PK, Thomas N, George S (2014) Glaucoma detection using artificial neural network. IACSIT Int J Eng Technol 6(2):158
3. Dey N, Roy AB, Das A, Choudari S (2012) Optical cup to disc ratio measurement for glaucoma diagnosis using harris corner. Eur J Sci Res 59. ISSN 1450-217X
4. Narasimhan K, Vijayarekha K, JogiNarayana KA, SivaPrasad P, Satish Kumar V (2011) Glaucoma detection from fundus image using opencv. Res J Appl Sci Eng Technol 62(5)
5. Ganesh Babu TR, Shenbagadevi S (2011) Automatic detection of glaucoma using fundus image. Eur J Sci Res. 59(1):22–32. ISSN 1450-216X
6. Zhang Z, Liu J, Cherian SN, Sun Y, Lim JH, Wong WK, Tan NM, Lu S, Li H, Wong TY (2009) Convex hull based optic cup ellipse optimization in glaucoma diagnosis. In: 31st annual international conference of the IEEE EMBS, Minneapolis, Minnesota, USA, 2–6 Sept 2009
7. Inoue N, Yanashima K, Magatani K, Kurihara T (2005) Development of a simple diagnostic method for the glaucoma using ocular fundus pictures. In: Proceedings of the 2005 IEEE engineering in medicine and biology 27th annual conference Shanghai, China, 1–4 Sept 2005
8. Li H, Chutatape O (2003) A model-based approach for automated feature extraction in fundus images. In: Proceedings of the ninth IEEE international conference on computer vision (ICCV 2003) 2-Volume Set 0-7695-1950- 4/03
9. High Resolution Fundus Image database. https://www5.cs.fau.de/research/data/fundusimages/
10. Opticdisc.org Database. http://www.opticdisc.org/library/normal-discs/page7.html

Enhanced Techniques to Secure Medical Image and Data Transit

S. Uma Maheswari and C. Vasanthanayaki

Abstract This paper displays the work related to the secured medical picture transmission in light of watermarking and encryption. Client particular watermark is installed into the LSB of unique picture. Implanting watermark in LSB does not influence the nature of picture. This watermarked picture is then encoded by utilizing a pixel repositioning calculation. Every pixel is repositioned in light of the rest of after division by number 10. This leftover portion lattice goes about as encryption key and is required at the season of unscrambling as well. Comprehensive analyses are completed on proposed approach. The outcomes demonstrate that the watermark inserted is intangible and can be effectively separated at the recipient. Likewise, the encoded picture has no visual importance with the first picture and histogram of scrambled picture is changed. Encoded picture can be decoded with no loss of data from the picture. From this decoded picture watermark can be removed which endures no misfortune in the watermark. PSNR esteems for an arrangement of medicinal pictures are fulfilling.

1 Introduction

This section talks about some fundamental ideas expected to comprehend whatever is left of the report. This part comprises of three areas: First segment portrays application space, i.e., Restorative imaging. Second segment depicts the idea of security and why we require security of restorative information. Third segment presents picture encryption and watermarking.

S. Uma Maheswari (✉)
Department of CSE, Kumaraguru College of Technology, Coimbatore, Tamil Nadu, India
e-mail: umamaheswari.s.cse@kct.ac.in

C. Vasanthanayaki
Department of ECE, Government College of Technology, Coimbatore, Tamil Nadu, India
e-mail: vasanthi@gct.ac.in

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_40

1.1 Medical Imaging

Medical Imaging is the Technique and Process of making visual portrayals of insides a body, for clinical examination and restorative mediation, and also visual portrayal of the elements of a few organs or tissues. Restorative imaging looks to uncover inner structures covered up by the skin and bones and also to analyze and treat maladies.

Restorative Disease Diagnosis is finished by three ways

1. Consultation—Information which is acquired from Patients.
2. Physical Examination—Inspection, Auscultation, Measurements.
3. Medical Tests—Laboratory Analysis, Bio-flag Analysis (ECG), Image Analysis.

Picture Analysis intends to separate significant data from pictures; for the most part from Digital Images. Essentially, an Image is a rectangular exhibit of pixels. It has a distinct stature and an unmistakable width checked in pixels. The tallness of a picture is the quantity of lines and the width of a picture is the quantity of segments in a cluster. In this manner the pixel exhibit is a lattice of $n * m$ where n shows number of lines and m demonstrates number of sections.

A computerized picture is a numeric portrayal of a two-dimensional picture. Contingent upon whether the picture determination is settled, it might be of vector or raster compose. Independent from anyone else, the expression "advanced picture" normally alludes to raster pictures or bitmapped pictures. There are three sorts of computerized picture which are twofold picture, shading picture, and dim scale picture.

1. **Double Images**: Binary pictures are advanced pictures that utilizations just 1-bit to speak to every pixel which may either be 1 or 0. II.
2. **Dark-scale pictures**: Grayscale pictures utilize 8-bits to speak to every pixel of a picture. Dim-scale pictures are alluded to as monochrome pictures. They contain just dim level data and no shading data.
3. **Shading pictures**: Color picture is a computerized picture that is comprised of hued pixels every one of which holds three hues comparing to Red, Green, and Blue. Shading pictures are additionally called as RGB Images.

1.2 Picture Security

In human services industry, patients' information are moving from paper records to electronic records. Electronic patient record is a PC-based memory that can be evaluated over system. Electronic patient record contains all the human services related data of a patient and joins a few undertaking-based electronic therapeutic records particular to a patient. Electronic patient records changed the method for putting away patients' data. Electronic social insurance frameworks empower well-being suppliers' to outline and actualize particular applications that can help specialists for conclusion and treatment of patients.

Electronic patient records have made chances to human services cheats including restorative data fraud. Erroneous medicinal records may prompt restorative abuse that will make destructive results a patient. Numerous studies accept that restorative data fraud is the same as money-related wholesale fraud, however, in reality medicinal wholesale fraud has all the more crushing consequences for patients for there is an absence of plan of action for patient to revise the false passages in their therapeutic records. In actuality, false restorative data can murder a patient.

With the appearance of electronic patient records, patients' information can without much of a stretch be shared among doctors, social insurance suppliers, attendants, supporting staff, restorative research, and general human services administrations. Essentially, most medicinal services data including patients' information are not created exclusively inside a doctor/tolerant relationship, however is produced from the assorted sources, for example, non-doctor authority, nurture professionals, general well-being officers, research centers, and other auxiliary social insurance callings. The sharing and in addition dissemination of patient data empowers beneficial therapeutic research, appropriate treatment of patients, and change in medicinal services quality.

Then again, persistent records represent a test to keep up data secrecy, honesty, and accessibility—the modernized record induces that the requester has a similar equipment and programming correspondence convention and in this way empowers simple access to information. This may open ways to the unapproved parties who may deceitfully take patients' information for individual advantages, modify patients' records, and uncover patients' therapeutic history. The high openness of patients' information has made it less demanding for culprits to attack patients' data privacy, respectability, and accessibility and submit medicinal services misrepresentation.

Need of information transmission through the system has prompted the need of more noteworthy security to ensure data against various assaults. Numerous procedures are utilized to give security to data in transmission, for example, Image Encryption, Digital Watermarking, Fingerprinting, Cryptography, Digital Signature and Steganography, and so forth.

1.3 Picture Encryption or Image Encryption

It is a procedure of changing picture into either a picture which is confused or into an information of other arrangement having no reference to the first picture. The info picture to an encryption calculation is by and large alluded as a plain picture and the yield encoded picture is alluded as a figure picture. The encryption is a reversible procedure in which the encoded picture can be unscrambled to recoup the first picture. This should be possible just by the general population knowing about the calculation and the key utilized as a part of calculation to scramble the picture.

Encryption should be possible utilizing either an open key of the client or a mystery key of client, and in view of the key sort it is named

I. **Public Key encryption**: where the encryption is finished utilizing the general population key of sender and the decoding is finished utilizing the mystery key of the beneficiary.
II. **Private Key encryption**: where the encryption is finished utilizing the private key of sender and the unscrambling is finished utilizing the mix of mystery key of the sender and people in general key of recipient.

Picture encryption can be utilized to ensure pictures very still, for example, pictures put away on PCs or capacity gadgets which are being uncovered through misfortune or burglary of PCs. Picture encryption as rest shields them from being revealed and shared.

Keeping in mind the end goal to encourage mystery correspondence, picture encryption has discovered noteworthy place very touchy territories, for example, military observation, satellite data frameworks, medicinal services, private video conferencing, and so forth. A portion of the calculations utilize customary content based encryption plans to encode pictures specifically, however it is not an attainable thought because of distinction in size of content and picture, and consequently expends more measure of time.

Encryption of picture can be either a visual encryption wherein the scrambled picture will be an aftereffect of a rearranging system connected on the first picture which changes all or a noteworthy number of pixel positions in the first picture or it may be a change based plan where the picture pixels esteems will be modified in an approach to shape another picture with very surprising arrangement of pixel esteems. In the primary plan, changing the places of pixels encodes the picture just outwardly yet the properties of picture stay same for instance estimate, determination, histogram and so on, then again, in second approach changing the power esteems infers the adjustment in histogram by which the character of the picture is covered. The encryption calculations can take inputs either in square mode or in stream of information bits. Contingent upon the sort of info they can likewise be named piece figure and stream figure.

1.4 Picture Watermarking

Digital picture watermarking innovation is a rising field in software engineering, cryptography, flag preparing, and interchanges. Computerized Watermarking is proposed, by its engineers, as the answer for the need to give proprietorship assurance to the picture.

Picture watermarking is the way toward inserting information called a watermark (otherwise called Digital Signature or Tag or Label) into a picture with the end goal that watermark can be identified or separated at recipient side to make a statement about the picture. A straightforward case of advanced watermark would be an unmistakable "seal" set over a picture to distinguish the copyright of the picture. However the watermark may contain extra data including the character of the proprietor or buyer of a specific duplicate of the picture.

Picture watermarks can be isolated in two composes in view of the deceivability of the watermark in the watermarked picture

1. **Visible watermark**: A procedure where the watermark is installed in the first picture in a way that the watermark is obvious to bare eyes. The came about picture, here, is the first picture with noticeably "stamped" watermark on it.
2. **Invisible watermark**: A procedure where the watermark is installed in a way that it is perceptually undetectable to human visual framework and can be recognized just utilizing the plan utilized for implanting it. The came about picture has no visual contrasts with the first picture.

Visible watermarks are by and large utilized as a part of uses where just the possession evidence should be added to demonstrate the wellspring of picture, for example, photography magazines where picture visual quality debasement is passable, then again, the imperceptible watermarks are utilized as a part of profoundly touchy applications, for example, military applications, restorative applications, mystery administrations, and so forth., where the substance of the picture are exceptionally essential and no corruption of the picture quality is normal.

Contingent upon the level of security the watermark can be inserted in two unique spaces of the picture viz.

1. **Spatial area**: Gives the position guide of pixels in the picture.
2. **Frequency area**: Gives the power estimations of pixels in the picture.

By and large spatial area watermarking is utilized when just the proprietorship points of interest of the picture should be sent at low security level, then again recurrence space watermarking is utilized as a part of utilizations where alter location is critical at recipient and furthermore the pictures are subjected to high dangers of assaults in transmission.

The key attributes are as given beneath

i. Difficult to see: The undetectable watermark ought not be detectable to the watchers nor should the watermark debase the nature of the substance of picture.
ii. Robustness: Watermarked picture ought to be powerful to assaults for the most part changes and lossy compressions and others like scaling, editing, pivoting, and so forth.
iii. Tamper protection: Watermark ought to be impervious to altering proposed to evacuate the watermark and recuperate unique picture, any such endeavors ought to be reflected from the recouped watermark. Minimum adjustment of pixels: It must not modify countless with a specific end goal to protect an adequate visual quality.

1.5 Encryption and Watermarking

Image encryption gives high level of security, yet the authentication of innovation cannot be accomplished from the encryption as it was. Scrambled picture is secured until the point when it is in the encoded frame yet after decoding it can be assaulted either to shape new picture or to control the substance of picture. Utilizing watermarking together with encryption can resolve this issue. The idea of encryption and watermarking intends to join encryption and watermarking with each other either in a solitary stage or at two unique stages. We could first watermark the picture and after that encode it, or first scramble the picture and afterward watermark the scrambled picture. Contingent upon the strategies utilized the first picture can be unscrambled at beneficiary and watermark is removed. Both the unscrambling and extraction can be connected freely at collector or should be performed in arrangement characterized by the connected plan.

1.6 Steganography

Steganography is the specialty of concealing messages inside unsuspicious medium. The reason for steganography is to conceal the presence of a message from an outsider. Cryptography is generally utilized with steganography. Steganography works in two phases, inserting then extricating. Amid the implanting stage, a key is utilized to insert a message in a cover medium bringing about a stego-question. The stego-protest is then transmitted along open channels to its goal. At the point when the stego-question is gotten, the inserted message is removed from stegoobject utilizing the known stego-key. Steganalysis is the craft of finding and rendering such undercover message. It will probably abstain from attracting doubt to the transmission of a concealed message. Steganalysis contains two principle issues: Detection and Distortion. Disadvantage of Steganography is if the key is known then the information can be recovered effortlessly. Consequently, no security is accommodated the encoded picture. The scrambled picture can be secured by joining this steganography with visual cryptography and advanced watermarking.

2 Related Work

Number of specialists has proposed a various encryption methods having diverse highlights of each for secure picture encryption reason. This section portrays the investigation of a portion of the accessible writing identified with picture encryption and watermarking.

2.1 Techniques Used for Picture Encryption

Procedures utilized for Image Encryption AES encryption calculation is open, which brings numerous issues for its security. So to take care of the well-being issues of AES encryption calculation, Yang et al. [1] proposed an approach which enhances AES encryption calculation by methods for utilizing disorder hypothesis. Results demonstrate that the attainability and rightness of the approach are sufficient. Patel and Ragha [2] have introduced a steganography strategy that utilizations ideas of picture preparing and wavelet change systems. This procedure gives a technique to concealing mystery paired information inside a cover picture without expanding the size or dynamic scope of the picture by methods for joining cryptography and steganography. Steganography when joined with encryption gives a secured method for mystery correspondence between two gatherings. Picture is encoded by utilizing a scrambling calculation. Higher PSNR is accomplished. Jyoti and Neginal [3] presented new secure picture transmission method which makes an important mosaic picture and furthermore changes the mystery picture into a secret fragment-obvious mosaic picture of a similar size and has an indistinguishable visual appearance from the objective picture. Determination of target picture is adaptable and does not require any database creation. Unique mystery picture can be separated almost entire from the got picture. Tasneem and Bhavni [4] utilized cryptography and steganography strategies together to build the security of the information while transmitting through systems. Proposed approach utilizes content information as watermark. Before installing content into the picture it is scrambled by utilizing AES calculation with key of size of length 8. The scrambled content is installed into picture utilizing Discrete Wavelet Transform (DWT) strategy and the resultant picture is transmitted to the recipient. Watermark installed is reversible and information can be decoded at recipient. Saraf et al. [5] proposed a way to deal with consolidate of C code, Code Composer Studio and DSP processor for content encryption. Content encryption utilizes 128-piece size of key and in addition plaintext. 16 strings of size 8 are framed for encryption. AES encryption calculation in CFB mode with PKCS5 Padding technique is utilized here for picture encryption. Maitri et al. [6] have presented Byte Rotation Algorithm for record encryption and decoding with minim delay. This calculation enhances the security and decreases time for record encryption and unscrambling. Aftereffects of Byte Rotation Algorithm are practically identical with those accomplished from AES encryption calculation. To enhance record security irregular key age of 128-bit is utilized. Khandelwal and Sahu [7] proposed a story picture steganography system to stow away both picture and key in shading spread picture utilizing Discrete Wavelet Transform (DWT) and innately coordinate grouping in view of Fuzzy C-Means Clustering. Results demonstrate that there is no visual similitude between stego-picture and unique picture and the data can be recovered totally. Singh and Singh [8] introduced idea which utilizes 64-bits Blowfish Algorithm, which is intended to expand security and to enhance execution. Calculation is utilized with variable key size up to 448 bits. It utilizes Feistel organize which repeats straightforward capacity 16 times. The blowfish calculation is protected against unapproved assault and runs quicker than the well-known existing calculations. Higher number of rounds makes the approach more secure. Since Blowfish has no known security feeble focuses so far it can be considered as a phenomenal standard encryption calculation. Banu and Velayutham [9] have broken down AES Encryption calculation to give adequate levels of security to ensuring the privacy of the information. A rest booking technique is investigated to diminish the postponement of alert telecom from any sensor hub in WSNs. In particular, here two decided activity ways are utilized for the transmission of caution message alongside level-by-level counterbalance based wake-up design. On a basic occasion a caution is immediately transmitted along one of the movement ways to a middle hub, and afterward it is instantly communicated by the inside hub along another way without crash. Nurhayati and Ahmad [10] completed research to plan a use of steganography utilizing Least Significant Bit in which the message is scrambled utilizing the Advanced Encryption Standard calculation (AES). The consequence of research demonstrates the steganography is required to conceal the mystery message, so the message is not anything but difficult to know other individuals who are not qualified. Tang et al. [11] accompanied another picture encryption and steganography plot. To start with, the mystery message is encoded through the mix of another dark esteem substitution task and position stage which makes the encryption framework solid. And after that the prepared mystery message is covered up in the cover picture to satisfy steganography. Trial comes about demonstrate that the plan proposed in this paper has a high security level and better picture stego-picture quality.

Rad and Hosseini [12] proposed Grid-Based Hyper Encryption application that uses the computational assets of numerous work area PCs keeping in mind the end goal to scramble/decode substantial information documents with a standout amongst the most effective and secure encryption calculations, the propelled encryption standard. Effortlessness of utilization and elite of GBHE settles on it a perfect decision for conveying in any association that needs this usefulness.

2.2 Procedures Utilized for Medical Image Watermarking

Balamurugan and Senthil [13] consolidated three distinctive research areas in particular unique finger impression biometric, cryptosystem, and reversible watermarking. Proposed framework utilizes the unique mark biometric for validation, symmetric and also open key for cryptography process for private information and reversible watermarking for honesty. To give CIA strategy to the MDBMS, compliance SMS is sent to the patient for the helpful, that their data had achieved securely to the comparing goal. Nemade and Kelkar [14] proposed a reversible watermarking strategy for hued restorative pictures utilizing Histogram moving technique the shaded pictures is changed over to YUV picture and watermark is implanted in the Luminance (Y) segment. Histogram move strategy is likewise assessed by applying the Technique on RGB parts. Kaur and Madanlal [15] proposed a visually impaired picture watermarking calculation which in light of both the Modified Fast Haar Wavelet Transform and the Redundant Second Generation Wavelet Packet Transform. The thought behind the proposed calculation is to break down the cover picture utilizing Modified Fast Haar Wavelet Transform and Redundant Second Generation Wavelet Packet Transform as indicated by the measure of the watermark. The watermark is installed in the fine-scale groups of the Redundant Second Generation Wavelet Packet Transform of the fine-scale groups of the last Modified Fast Haar Wavelet Transform deterioration level of the host picture. Paul and Sunitha [16] proposed another technique for expanding the limit and protection of watermarking in light of round histogram tweak by taking the host framework as shading picture. By expanding the limit it can be helpful for secure transmission of tremendous databases like healing facility records through the system. Anandkumar and Mukeshgupta [17] created two calculations for implanting and extraction process. The first picture is first DWT-decayed, and after that the shading watermark is embedded. Watermark implanting is connected in various recurrence groups of the picture and PSNR and NC plots are drawn for all the recurrence groups. Bolandt and Kuannitlh [18] proposed watermarking plan in light of square DCT WT- and FFT-based calculation. The approach watermarks are intended to be imperceptible even to a cautious spectator yet contain adequate data to interestingly distinguish both the root and planned beneficiary of a picture with a low likelihood of blunder. The further improvement of powerful mistake revision codes and computerized signature methods is to be finished. Another approach of watermarking is proposed by Chang and Lu (2008) [19] which depend on installing watermark into dim level pictures as per pixel connection between them. The creator guarantees the way to deal with be lossless as far as nature of de-watermarked picture. Approach is fit for concealing expansive measure of data and can reestablish the first picture with no misfortune, and this is appeared by comes about. Likewise it has been guaranteed that the plan beats RS-installing plan and Tian's plan and consequently can essentially be connected. Shukla [20] proposed continuous copyright security calculation utilizing both obvious and undetectable watermarking plans and furthermore execution of constant picture handling procedures on Android and Embedded Platform. The idea depends on DCT and OpenCV. Pictures caught from the Smart-telephone camera are effectively watermarked utilizing this framework. The creators guarantee that DCT gives a productive strategy to executing imperceptible watermarking process. Obvious watermarking includes inserting content watermark which is strong to basic picture handling activities. Mathon et al. [21] utilized bending improvement for secure spread range watermarking for grayscale pictures. This approach of contortion minimization depends on components of transportation hypothesis and henceforth claims to accomplish solid security properties and is called as Transportation Natural Watermarking. Here, the multiresolution picture is disintegrated and joined with a multiplicative installing at dispersion level. Installing contortion and its visual effect on picture can be decreased utilizing the proposed approach and this can be found in comes about.

Peng [22] proposed a square based particular esteem decay of the picture and the discrete system strategy to install and separate watermark to and from the advanced picture. It has tackled particular esteem disintegration issue of poor heartiness. The idea depends on the neural system and solitary esteem disintegration. The creator guarantees well extraction all through the normal picture preparing and packing tasks. Kamran and Farooq [23] proposed an approach of watermarking giving great nature of picture after watermark extraction. Creator utilized component extraction and positioning to watermark pictures. The utilization of highlight positioning adds to safeguarding of data contained inside the picture. To make and implant watermark to unique picture, Particle Swarm Optimization Algorithm is utilized. The proposed plot is guaranteed to be flexible to an assortment of assaults and is affirmed by comes about demonstrated Kishor and Vankat [24] proposed a RSA-DWT-based restorative picture watermarking. Open key encryption of watermark is done before inserting watermark into the picture. This encoded watermark is then implanted into the picture utilizing DWT. Results demonstrate agreeable estimations of PSNR and zero ability to see of watermark. Dragoi and Coltuc [25] proposed a watermarking plan which depends on neighborhood expectation to accomplish contrast extension reversible watermarking. Here, a minimum square indicator is figured for every pixel on a square piece, focused on the pixel and the expectation blunder relating to that pixel is extended. This expectation mistake is then recouped at beneficiary with no extra data. Forecast setting relies upon the span of squares utilized as a part of extension. Fitting square sizes for every setting are resolved and are given in comes about. Results got by Rhombus setting are the best among the various settings. Su et al. [26] presents an approach of utilizing Scale Invariant Feature Transform and broadened pilot signals for watermarking. This approach tends to the issue of synchronization in watermark discovery at beneficiary. The watermarking signal is inserted in the invariant locales which are shaped by applying interest point extraction as scale invariant highlight change. Square frameworks are utilized to install flag. This approach accomplishes right harmony between effectiveness of flag recognition and heartiness of watermark. Installed flag are guided by pilot flag contributing towards synchronization in discovery. At discovery when a similar intrigue point is separated, the related relative parameters of matrices are acclimated to distinguish concealed pilot flag accomplishing synchronization and effective recovery of watermark and unique picture. Panyavaraporn [27] concocted an approach of utilizing wavelet change to watermark QR code pictures. The parallel watermark is implanted into a chosen sub-band which can be both of LH, HL, or HH sub-groups. The outcomes demonstrate that the plan can withstand fundamental assaults acceptably yet is not vigorous against assaults, for example, solid clamor, high-pressure geometric change and impediment and thus restricts the execution of the plan. Abdallah et al. [28] presented wavelet-based watermarking which does not require unique picture or any pilot picture to recreate watermark at decoder. The watermark is installed into the coarsest scale wavelet coefficient. Decay up to level 3 and watermark of size equivalent to sub-band is utilized. Plan utilizes quantization of wavelet coefficients in parallel way. Unlike customary wavelet-based strategies, this technique has less corruption.

Lin and Wu [29] proposed a way to deal with watermark 3D pictures. Here numerous watermarking is utilized to take care of substance insurance issue of Depth Image Based Rendering (DIBR) 3D pictures. A 3D picture contains three pictures in itself viz. left-eye picture, right-eye picture, and a middle picture. The inside picture and profundity picture created by content supplier is called as (DIBR) 3D picture. Both the, common orthogonality and request of implanting assumes imperative part watermarking 3D pictures. The upside of the approach is that it need not bother with unique information and profundity picture amid watermark discovery. Results demonstrate that the approach is secure to pressure and commotion assaults and furthermore does not get influenced by run varieties and standard separation modification up to certain degree. Zhang et al. [30] did examine for watermarking on Perceptual Quality Metric in view of second order insights. Abusing the perceptual twisting with human recognition is key for strong watermarking plan.

The proposed plan of SOS performs superior to some cutting edge measurements and connects with a few databases of pictures by methods for surface concealing impact and difference affectability in Karhunen–Loeve change space. Utilization of straightforward metric guarantees fat usage and results demonstrate that the heartiness is made strides. Creator likewise proposes association of outsider measurements to rate the nature of watermarked picture. Tang and Hang [31] proposed utilization of Mexican Hat wavelet scale communication strategy in light of highlight focuses implanting and extraction as watermarks.

This is accomplished utilizing picture standardization which is connected to non-covered pictures independently. A succession of 16-bit signals is implanted in unique picture enhancing strength of watermarking. Reference picture is not required at indicator to extricate watermark. The synchronization issue is overwhelmed by utilizing visual point on picture and invariance focuses lessen the watermark look space. Guerrini et al. [32] proposed a strategy to watermark High Dynamic Range (HDR) pictures having high luminance esteems. The HDR pictures are tone mapped utilizing TM calculation and after that watermarked utilizing Quantization Index Modulation (QIM). Watermarking framework produced for LDR pictures is utilized here in LogLuv space. The discovery of watermark relies upon the TM calculation. Weakness is that the plan isn't completely visually impaired and choice of squares to insert information isn't conceivable.

2.3 Strategies Utilized for Image Security

Sundari et al. [33] thought of three-method watermarking, steganography and cryptography to give abnormal state of security. To begin with the cover picture is be packed utilizing JPEG pressure calculation, then the message to be sent is encoded utilizing RSA encryption calculation and later the scrambled message bits and the bits of watermark are installed into cover picture. Patel and Patel [34] proposed consolidated technique of cryptography, steganography and computerized watermarking to stow away secure picture with watermark logo inside cover picture. For this reason they utilized DCT, DWT, SVD and RSA approach. Utilizing DCT, scrambled watermark logo (encryption performed utilizing RSA) is covered up inside secure picture bringing about Stego-picture. Gayathri and Nagarajan [35] presented approach with data stowing away in picture utilizing Zig-Zag filtering design which is more mind boggling calculation in Steganography again encoded as offers.

The offer is implanted into the host picture utilizing Least Significant Bit Insertion Technique (LSB). The plan gives more secure and significant mystery shares that are strong against various assaults. Kishor et al. [36] proposed RSA-DWT-based therapeutic picture watermarking like MRI. Watermark picture is encoded with key created RSA calculation. Scrambled patient picture is utilized as payload which is installed into Medical picture in Wavelet space. Exploratory result demonstrates that RSA-DWT shows predominant security on unsecured network. Suganya et al. [37] proposed another joint watermarking and encryption framework which ensures from the earlier and a posteriori assurance of medicinal pictures. It blends the stream figure calculation or piece figure calculation with the QIM-based watermarking system. The proposed strategy utilizes two encryption calculations in particular RC4, AES are Stream figure, Block figure calculation separately. The framework offers access to embed two messages in the spatial area and encoded space individually amid encryption process. These two messages are utilized to confirm the unwavering quality of pictures in decoding part. Li and Bai [38] built up a calculation with computerized watermarking plan which utilizes a piece of sign succession of DCT coefficients as a component of vector pictures, at same time watermark is encoded by strategic guide to upgrade classification.

3 Proposed Methodology

This part gives the portrayal about the approach we have proposed. It comprises of the means engaged with proposed approach and the stream of the approach alongside information stream in the framework. Stream outlines for the approach are likewise recorded here.

Advancement in the web, today, has improved data sharing over the more extensive region. Both the visual and printed data of a patient can undoubtedly be transmitted the whole way across the world. In any case, the simplicity of data trade has presented security issues for the data being transmitted over the web, as the data can undoubtedly be stolen and wrongfully altered at any phase without having any personality of the individual.

Pictures being shared over the web can be ensured by utilizing some safety efforts like encryption and watermarking. Work proposed here spotlights on the mean to give security to pictures being shared over the web. The approach utilizes watermarking and encryption both to give confirmation information and visual security to restorative pictures. Watermark implanting is finished utilizing LSB substitution and encryption of watermarked picture and patients' information is encoded utilizing scrambling technique. The approach is as examined beneath.

The proposed work can be isolated into four stages as tail

i. Installing watermark in the first picture.
ii. Encryption of Watermarked picture.
iii. Encryption of patients' data.
iv. Decoding of patients' picture and extraction of watermark.
v. Decoding of patients' data.

In first stage, the watermark is implanted into unique picture to shape a watermarked picture. Both the first picture and watermark are utilized as a part of the picture arrange. Information from other arrangement with the exception of picture cannot be inserted in the first picture.

In second stage, watermarked picture in scrambled utilizing a rearranging calculation talked about later. An encryption is likewise produced in this same stage which is utilized for decoding moreover.

At that point in third stage the data of patient is encoded utilizing Byte Replacement Encryption Algorithm (BREA). All characters from the patients data graph are changed over encoded utilizing a key framed utilizing arbitrary dissemination of numbers.

At last at the recipient, got encoded watermarked picture is unscrambled utilizing the same keys utilized for encryption first and the first watermark is separated from the decoded picture. Removed watermark can be utilized to identify the legitimacy and innovation of picture. Implanting Watermark Watermarking is the way toward concealing a mystery picture or message into the first picture so as to give genuineness to the picture. Message picture can be installed in the first picture utilizing pixel substitution strategy wherein one of the bit planes of the first picture is supplanted utilizing the message picture. Pixels in the picture have 8-bit esteem speaking to their shading force. In the event that the Least Significant Bit (LSB) of these 8-bits is modified the aggregate estimation of the pixel will change just by 1. For instance, changing LSB of pixel with esteem 128 from 1 to 0 will change the incentive to 129 and outwardly a human eye can't recognize such a little distinction.

Watermark, in the proposed approach is of paired frame and is inserted by methods for supplanting the LSBs of the first picture. To start with the first LSB plane of the picture is made 0, at that point the double watermark is added to this LSB plane. LSB watermarking is one of the most straightforward strategies used to watermark pictures. It is a result of the low intricacy and high recuperation proficiency of the watermark. Watermark can simply be recuperated with no misfortune from a protected picture.

3.1 Encryption of Watermarked Picture

Watermarked picture from stage one is scrambled in this second stage. Point of encoding picture is to secure picture against stealth and control assaults. This is on account of even a slight control in the patients' therapeutic picture can cause him his life. Encryption is the way toward making an interpretation of picture into incoherent data spoke to in either picture frame or a content shape.

Approach proposed in this work utilizes disintegration of pixel esteems with a specific end goal to rearrange their positions. Encryption process, proposed here, rearranges every one of the pixels in the first picture and furthermore changes their unique pixel force esteems with a specific end goal to adjust the histogram of the last encoded picture.

Encryption calculation can be partitioned into a few phases. To reposition a pixel rest of the division of unique incentive by 10 is utilized. Not the genuine esteem but rather the remainder of division is set at the position indicated by the rest of.

Stage 1: In initial step unique pixel esteem is partitioned by 10 and both leftover portion and remainder are recorded in frameworks *Q* and *R* individually.
Stage 2: another lattice with same number of lines as that of unique picture and with sections 10 times as that of the first picture is framed, to such an extent that every one of the components take arbitrary numeric esteems from 0 to 255.
Stage 3: First component in the *Q* is set at position spoke to by first component of R, in new lattice. Second component from *Q* is put at the position spoke to by second element+10 of *R*.
Stage 4: Perform stage 3 for all components in first line of *Q*.
Stage 5: Repeat stage 4 for all sections of *Q*.
Stage 6: Finally change over every one of the components from scrambled lattice into their ASCII structures to accomplish a content document with encoded picture information. Components of this content document are then mapped in the scope of 0–255 by methods for utilizing their separate ASCII codes to get the last encoded picture. Here, the network with leftover portion esteems, i.e., *R* goes about as a key for encryption and is required to decode the scrambled picture. Figure demonstrates the picture encryption process in detail.

3.2 Encryption of Patients' Information

Patient information is contribution to the type of content. This content information is encoded utilizing Byte Replacement Algorithm. BREA calculation encodes content information examined in squares and each piece is then scrambled in parallel way. A key is produced for encryption which is additionally utilized for unscrambling moreover. BREA tests input information into strings of size 16 characters. Each such string is then reshaped into a lattice of size 4 * 4. Every one of the components are then supplanted by their equal ASCII codes. Another grid, a key framework, is produced by methods for haphazardly choosing all components remarkably from 1 to 26. Components of key network are then scaled by methods for taking mod by 2. This key network is then included with each square component by component. At that point, first component of first line is turned toward the finish of column by moving all components to left. Initial two components of second column are pivoted by moving different components towards left lastly initial three components of third line are turned toward the end. Fourth line is kept unaltered. Likewise, first component of first section is pivoted toward the end by moving different components towards up. Initial two components of second section are turned toward the end lastly initial three components from third segment are pivoted towards drawback. Fourth segment is kept unaltered.

4 Results and Discussions

Following are the depictions of the outcomes for the proposed approach. Figure 1 speaks to the real literary information entered by the patient or an agent, encoded information accomplished from the Byte Replacement Encryption Algorithm and the decoded data acquired utilizing reverse BREA.

From Fig. 2, the data entered was "Abc Def Xyz 123 789", a blended string of capitalized letters, bring down case letters, clear spaces and numbers. Encoded string accomplished is g "2bX2c z4!Dz!Be 8 9 8". It can be obviously observed that encoded does not look like with the first content specifically and is difficult to get it. It additionally does not have any standard significance.

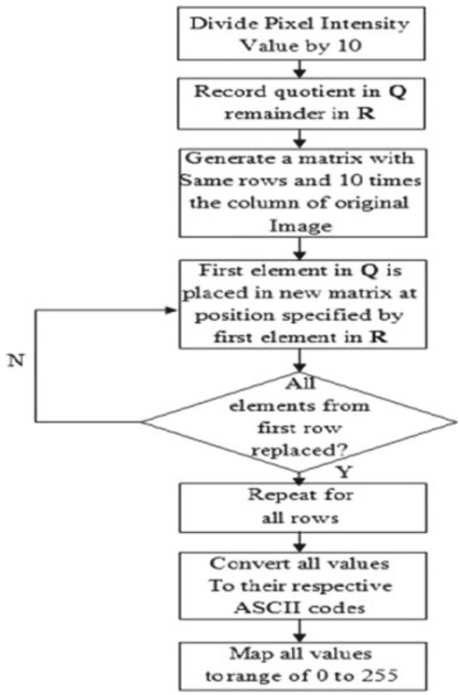


Fig. 1 Encryption of watermarked image flow diagram

Fig. 2 Original, encrypted and decrypted textual information

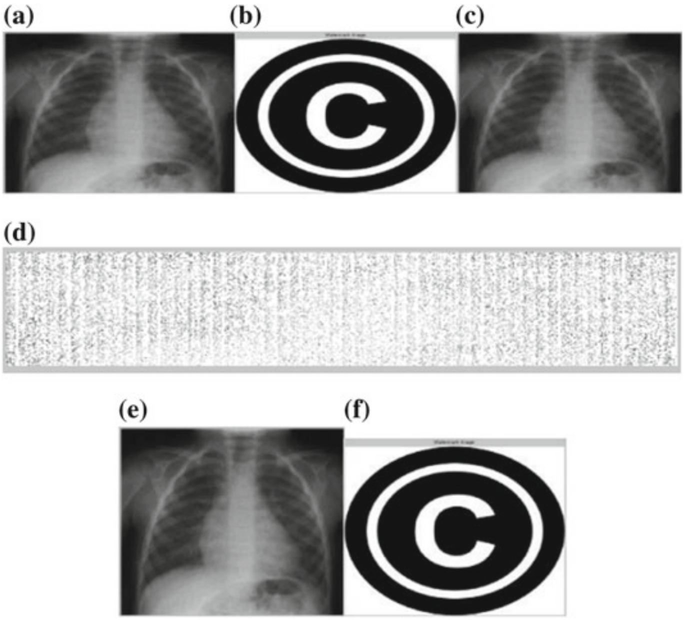


Fig. 3 **a**–**f**: Original X-Ray image, watermark image, watermarked image, encrypted image, decrypted image and extracted watermark

Unique X-Ray picture, watermark picture, watermarked picture, encoded picture, unscrambled picture and separated watermark Fig. 3 demonstrates the aftereffect of finish procedure of encryption and watermarking on picture alongside decoding and extraction comes about. It can be seen from figure (c), (d) and (e) that watermen picture can be scrambled without leaving any perceptual connection with unique picture. Likewise the expanded size presents bigger number of conceivable outcomes for false pixel repositioning decreasing the assault probability and the same scrambled picture can be decoded with no misfortune.

Again on account of the altogether changed pixel esteems histogram of unique picture does not co-relate with encoded picture PSNR Analysis Higher PSNR esteems

Table 1 PSNR analysis of images

Image	PSNR
USG image	33.5732
X-Ray image	33.5716
Brain image	34.5979
MRI image	33.6976
Kidney image	34.6006

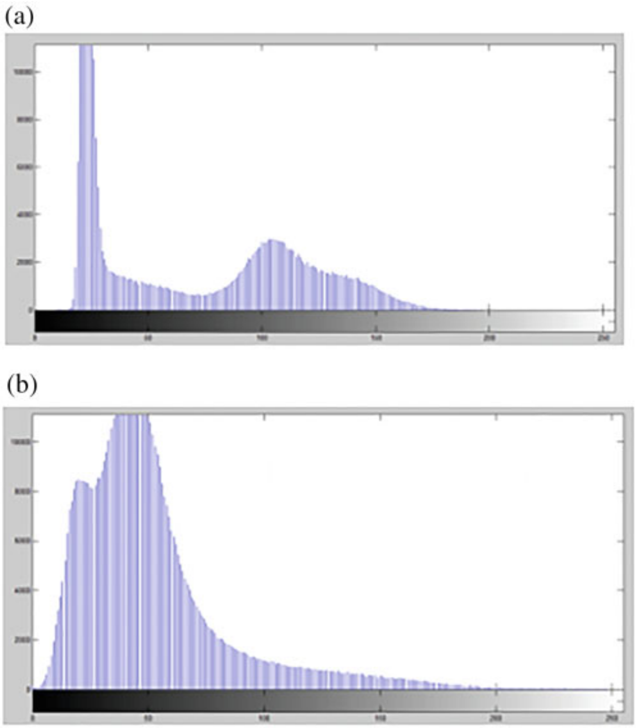


Fig. 4 **a**, **b** Histogram of original Sonography image and histogram of encrypted Sonography image

are wanted for better outcomes. Table 1 shows estimations of PSNR for Rib Cage Image when experienced Salt ‘n’ Pepper assault.

It can be plainly noted from Fig. 4 that encryption makes the picture subtle as well as changes the histogram which decreases the odds of histogram assaults, or endeavors of recovering picture from histogram.

5 Conclusion

In proposed plot watermarking of picture is done before encryption. Due to earlier watermarking the encryption is given more concentration in the approach. We have utilized LSB watermarking procedure, a simple to execute and viable, for inserting the watermark into the cover picture. Furthermore, to encrypt the picture we have a repositioning calculation which rearranges the pixels in view of the aftereffect of a division. It is then comprehended from the outcomes that the proposed approach has prevailing to watermark and scramble picture and the recovery of picture and watermark is likewise conceivable with no misfortune. An important point is that the unscrambled picture keeps up high visual quality. Additionally, the removed watermark is similar to the first one inserted and consequently can be utilized for validation and copyright assurance.

References

1. Yang Z-H, Li A-H, Yu L-L, Kang S-J, Han M-J, Ding Q (2015) An improved AES encryption algorithm based on chaos theory in wireless communication networks. In: Third international conference on robot, vision and signal processing
2. Patel K, Ragha L (2015) Binary image steganography in wavelet domain. In: International conference on industrial instrumentation and control (ICIC)
3. Jyoti RH, Neginal J (2015) A new secure image transmission technique via secret-fragment-visible mosaic image. Int J Adv Res Comput Eng Technol (IJARCET) 4(5)
4. Tasneem SF, DurgaBhvni S (2014) Secure data transmission using cryptography and steganography. Int J Emerg Trends Electr Electron 10(9)
5. Saraf KR, Jagtap VP, Mishra AK (2014) Text and image encryption decryption using advanced encryption standard. Int J Emerg Trends Technol Comput Sci (IJETTCS) 3(3)
6. Maitri PV, Sarawade RV, Deokate ST, Patil MP (2014) Secure file transmission using byte rotation algorithm in network security. In: International conference for convergence of technology
7. Khandelwal P, Sahu B (2014) Novel technique data-hiding scheme for digital image
8. Singh P, Singh K (2013) Image encryption and decryption using blowfish algorithm in MATLAB. Int J Sci Eng Res 4(7)
9. Jesima Banu A, Velayutham R (2013) Secure communication in wireless sensor networks using AES algorithm with delay efficient sleep scheduling. In: 2013 IEEE international conference on emerging trends in computing, communication and nanotechnology (ICECCN 2013)
10. Buslim N, Ahmad SS Steganography for inserting message on digital image using least significant bit and AES cryptographic algorithm
11. Tang H, Jin G, Wu C, Song P (2009) A new image encryption and steganography scheme. In: 2009 international conference on computer and communications security
12. Rad NB, Shah-Hosseini H (2008) GBHE: Grid-based cryptography with AES algorithm. In: 2008 international conference on computer and electrical engineering
13. Balamurugan G, Senthil M (2016) A fingerprint based reversible watermarking system for the security of medical information
14. Nemade H, Kelkar V (2016) Reversible watermarking for colored medical image using histrogram shifting method
15. Kaur S, Lal M (2015) An invisible watermarking scheme based on Modified Fast Haar Wavelet Transform and RSGWPT. IEEE

16. Paul A, Sunitha EV (2015) Distortion less watermarking of relational databases based on circular histogram modulation. In: 2015 international conference on circuit, power and computing technologies
17. Kumar A, Gupta M (2015) Semi visible watermarking scheme based on DWT and PCA. IEEE
18. Boland FM, Kuannitlh JJ (2007) Watermarking digital images for copyright protection. Image Processing and Its Applications, 4–6 July 2007
19. Chang C-C, Lu T-C (2007) Noise features for image tampering detection and steganalysis. IEEE
20. Shukla RJ (2014) Platform independent real time copyright protection embedding and extraction algorithms on android and embedded framework. IEEE
21. Mathon B et al (2011) A parametric solution for optimal overlapped block motion compensation. IEEE
22. Peng J (2011) The research on digital watermarking algorithm based on neural networks and singular value decomposition. IEEE
23. Kamran M, Farooq M (2009) A new spatial decomposition scheme for image content-based watermarking. IEEE
24. Kishor PP, Yankat N (2014) Medical image watermarking using RSA encryption in wavelet domain. IEEE
25. Dragoi I-C, Coltuc D (2014) Local-prediction-based difference expansion reversible watermarking. IEEE Trans Image Process 23:1779–1790
26. Su P-C, Wu C-Y et al (2013) Geometrically resilient digital image watermarking by using interest point extraction and extended pilot signals. IEEE Trans Inf Forensics Secur 8:1897–1908
27. Panyavaraporn J (2013) QR code watermarking algorithm based on wavelet Symposium transform. In: 2013 13th international symposium on communications and information technologies (ISCIT), Sept 2013
28. Abdallah HA, Hadhoud MM, Abdalla Hameed (2011) Blind wavelet based image watermarking. Int J Signal Process Image Process Patter Recogn 4(1)
29. Lin Y-H, Wu J-L (2011) A digital blind watermarking for depth-image-based rendering 3D images. IEEE Trans Broadcast 57:602–611
30. Zhang F et al (2011) Spread spectrum image watermarking based on perceptual quality metric. IEEE Trans Image Process 20:3207–3218
31. Tang C-W, Hang H-M (2003) A feature-based robust digital image watermarking scheme. IEEE Trans Signal Process 51:950–959
32. Guerrini F, Okuda M, Adami N, Leonardi R (2011) High dynamic range image watermarking robust against tone-mapping operators. IEEE Trans Inf Forensics Secur 6:283–295
33. Sundari M, Revathi PB, Sumesh S (2015) Secure communication using digital watermarking with encrypted text hidden into an image
34. Patel P, Patel Y (2015) Secure and authentic DCT image steganography through DWT—SVD based Digital watermarking with RSA encryption. In: 2015 fifth international conference on communication systems and network technologies
35. Gayathri R, Nagarajan V (2015) Secure data hiding using Steganographic technique with Visual Cryptography and Watermarking Scheme. In: IEEE ICCSP 2015 conference
36. Kishor PV, Venkatraman N, Reddy SS (2014) Medical image watermarking using RSA encryption in wavelet domain. IEEE
37. Suganya G, Amudha K (2013) Medical image integrity control using join encryption and watermarking techniques. IEEE
38. Dong L, Bai H (2012) Medical image watermarking algorithm with encryption by DCT and logistic. IEEE
39. Farah T, Hermassi H, Rhoouma R, Belghith S (2012) Watermarking and encryption scheme to secure multimedia information. IEEE
40. Khan MI, Jeoti V, Malik AS, Khan MF (2011) A joint watermarking and encryption scheme for DCT based codecs. In: 2011 17th Asia-Pacific conference on communications (APCC)
41. Bouslimi D, Coatrieux G, Roux C (2011) A joint watermarking/encryption algorithm for verifying medical image integrity and authenticity in both encrypted and spatial domains

A Spectral Approach for Segmentation and Deformation Estimation in Point Cloud Using Shape Descriptors

Jajula Kalyani, Karthikeyan Vaiapury and Latha Parameswaran

Abstract In this paper, we propose a new framework for segmentation and deformation estimation in texture-less point clouds. Given a reference point cloud and a corresponding deformed point cloud, our approach first segments both the point clouds using OBB-LBS (Oriented Bounding Box-Laplace Beltrami Spectral) and estimates the semi-global dense spectral shape descriptors. These coarse descriptors identify the segments which need to be further investigated for localizing the area of deformation at a finer level.

1 Introduction and Related Work

Deformation detection and localization are quite useful in many real-world applications such as for quality check, maintenance of aircrafts, automobiles, town planning in smart cities, and surgery in medical domain. For instance, if a 3D model of an object is given and in another if the scan of the object is deformed in some locations, then detecting and estimating that change can lead to a fast quality inspection.

In this proposed approach we have used Laplace Beltrami Spectral (LBS) signatures as they are invariant to isometry. Isomerism exists in 3D models when there are similar structures on either side and it has been widely studied. In this work, we leverage a combination of OBB estimation and quantum inspired algorithms such as spectral wave kernel signatures to estimate the deformation.

J. Kalyani · K. Vaiapury
TCS Innovation Labs, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Coimbatore, India
e-mail: cb.en.p2cvi15002@cb.students.amrita.edu

K. Vaiapury
e-mail: karthikeyan.vaiapury@tcs.com

J. Kalyani · L. Parameswaran (✉)
Department of Computer Science and Engineering, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Coimbatore, India
e-mail: p_latha@cb.amrita.edu

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_41

Various algorithms are explained for change detection in 3D and 2D space in the literature. A systematic survey of change detection algorithms of 2D space like pixel-based-, object-based-, shading model, predictive models, and hypothesis testing are explored in [1, 2] and for 3D space methods like detecting a new object in the scene and detection of change existing model are explored in [3–6]. The demerits in 2D space like illumination, temporal, and spatial constraints, shadows, atmospheric conditions can be excluded if the change detection is done in 3D space. The challenges in 3D space are due to registration errors, density difference between target and source point clouds, irregularities in the point cloud, sensitivity of the detection method to small changes and sensor noise. Various domains such as aircraft [7], building [8] and medical [9] are explored for change detection in 3D. In [10] the authors have extracted Gabor wavelet-based features for face recognition. These features can also even be extended to detecting changes in images. In [11] the authors have used Independent Component Analysis (ICA) features for authenticating images by watermarking. A similar approach can be tried to extract block-wise changes.

To handle computational complexity tradeoff, in [12], a new oriented bounding box (OBB) has been proposed for extracting regional area descriptor and has been used for 3D point cloud registration when no coordinate references exist. These techniques can only capture global surface properties and cannot handle dataset with isometric or topological changes which spectral signatures can capture in a succinct manner. In this article, we propose a OBB and LBS based approach for detecting changes in 3D point cloud.

2 Proposed Work—Deformation Detection and Localization Framework

The proposed framework given in Fig. 1 consists of three key components. In the first component, the estimation of OBB is done on both individual point cloud of source and target data using covariance-based method and is partitioned into segments via sub-blocks. The salient partitioned blocks are fed to the second component to get the LBS spectral signatures. These signatures are fed into the third component where the interpolation of signatures is done to estimate the deformation and visualization of point cloud with deformation.

A. *Estimation of OBB and Partitioning*

This proposed approach involves estimation of OBB for both source M and target point cloud M' analogous to Chen et al. [12] based on covariance-based method and partitioning of cloud into sub-blocks.

The estimation of OBB is done as given below

1. Let p represent the number of vertices in a point cloud given as $X = (a, b, c)$
2. Calculate centroid of the point cloud

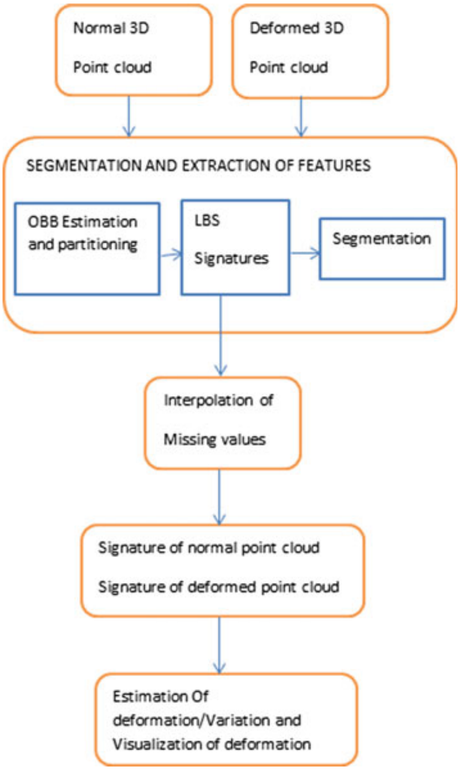


Fig. 1 Framework for deformation estimation and segmentation

$$\mu = (\mu_a, \mu_b, \mu_c)$$
$$\mu_a = \sum_{j=0}^{p} a_j \mu_b = \sum_{j=0}^{p} b_j \mu_c = \sum_{j=0}^{p} c_j$$

3. Calculate the covariance matrix for the point cloud

$$\text{Covariance} = \frac{1}{p} X(X)^T$$

4. Calculate eigenvalues and eigenvectors for covariance matrix and sort them as $\lambda_1 \geq \lambda_2 \geq \lambda_3$ with their respective eigenvectors $v_1 \geq v_2 \geq v_3$
5. OBB represented by $c = (a_c, b_c, c_c)$ and cc_1, cc_2, cc_3 is calculated using the following equations

$$a'_j = v_{11}(a_j - \mu_a) + v_{12}(b_j - \mu_b) + v_{13}(c_j - \mu_c)$$
$$b'_j = v_{21}(a_j - \mu_a) + v_{22}(b_j - \mu_b) + v_{23}(c_j - \mu_c)$$
$$c'_j = v_{31}(a_j - \mu_a) + v_{32}(b_j - \mu_b) + v_{33}(c_j - \mu_c)$$

$$a'_{\min} = \min\left\{a'_j, \forall j = 1, 2, \ldots p\right\}$$
$$b'_{\min} = \min\left\{b'_j, \forall j = 1, 2, \ldots p\right\}$$
$$c'_{\min} = \min\left\{c'_j, \forall j = 1, 2, \ldots p\right\}$$
$$a'_{\max} = \max\left\{a'_j, \forall j = 1, 2, \ldots p\right\}$$
$$b'_{\max} = \max\left\{b'_j, \forall j = 1, 2, \ldots p\right\}$$
$$c'_{\max} = \max\left\{c'_j, \forall j = 1, 2, \ldots p\right\}$$

6. OBB parameters are computed as:

$$a_c = \mu_a + v_{11}a'_{\min} + v_{21}b'_{\min} + v_{31}c'_{\min}$$
$$b_c = \mu_b + v_{12}a'_{\min} + v_{22}b'_{\min} + v_{32}c'_{\min}$$
$$c_c = \mu_c + v_{13}a'_{\min} + v_{23}b'_{\min} + v_{33}c'_{\min}$$
$$cc_1 = (a'_{\max} - a'_{\min})v_1 \quad cc_2 = (b'_{\max} - b'_{\min})v_2$$
$$cc_3 = (c'_{\max} - c'_{\min})v_3$$

7. Finding corners from the above parameters is tricky $c_1 = cc_1 - c$ and based on dataset it may be $c_1 = cc_2$

$$d_1 = cc_{1a} \sim \mu_a$$
$$d_2 = cc_{1b} \sim \mu_b$$
$$d_3 = cc_{1c} \sim \mu_c$$

8. Corners are computed using

$$c_2 = (c_{1a}, \mu_b + d_2, c_{1c})$$
$$c_3 = (c_{1a}, c_b, \mu_c + d_3)$$
$$c_4 = (c_{1a}, \mu_b + d_2, c_{1c} + d_3)$$
$$c_5 = (\mu_a + d_1, c_{1b}, c_{1c})$$
$$c_6 = (\mu_a + d_1, \mu_b + d_2, c_{1c})$$
$$c_7 = (\mu_a + d_1, c_{1b}, c_{1c} + d_3)$$
$$c_8 = (\mu_a + d_1, \mu_b + d_2, c_{1c} + d_3)$$

3 Experimental Results

This proposed algorithm has been implemented on sample point clouds for experimentation. Using the above steps, OBB corners for a sample cloud of aircraft and the results are shown in Fig. 2.

Let the corners of a sample cloud be (0, 0, −200), (2000, 0, −200), (2000, 0, 200), (0, 0, 200), (0, 1000, −200), (2000, 1000, −200), (0, 1000, 200), (2000, 1000, 200); then the partition of cloud can be done by calculating the midpoints of corners and the sub-blocks of partition is shown in Table 1. The horizontal partition of the cloud is done by taking the midpoint of $z_{\min}$ and $z_{\max}$ of aircraft corners.

After obtaining the OBB corners of aircraft from the covariance-based method, the midpoints of corners for partition are calculated and with the reference of Table 1, the horizontal and vertical partition is done and the results are shown in Fig. 3.

B. *Laplace Beltrami Spectral Analysis*

The divergence of the gradient of manifold ***f*** is known as Laplace Beltrami [13]. A manifold is known as a topological space which is locally Euclidean. Generally, any object which is as “flat” as possible when a small space is taken is called as manifold. The Eigen function of Laplace Beltrami differential operator captures the surface global properties. In the beginning, for shape descriptors the Eigen values of this operator were used but as they are unable to differentiate iso-spectral shapes a new signature has been proposed based on nodal counts of Eigen functions. The cropped set of eigenvalues of Laplace Beltrami is called as “Shape-DNA [14]”. Potential applications of these Eigen functions, as exemplified include signal processing on surfaces, and geometry. The authors in [16] and [17] have discussed Laplace

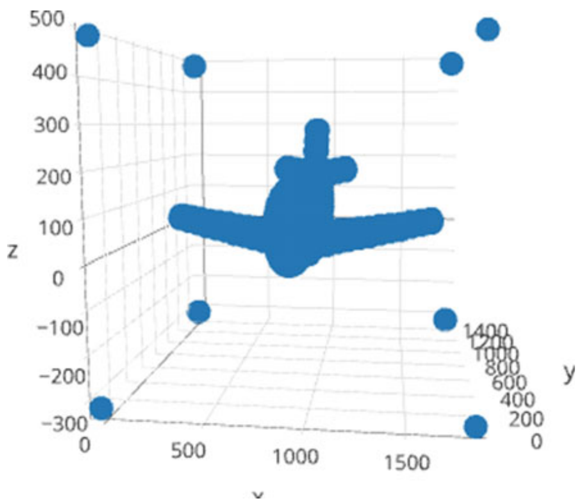


Fig. 2 OBB corners with aircraft cloud

Table 1 Partition block table sample H—Horizontal, V—Vertical, Q—Quad and O—Oct B—Block [12]

Block	x_{min}	x_{max}	y_{min}	y_{max}	z_{min}	z_{max}
HB_1	0	2000	0	1000	0	200
HB_2	0	2000	0	1000	−200	0
VB_1	0	1000	0	1000	−200	200
VB_2	1000	2000	0	1000	−200	200
QB_1	0	1000	0	1000	0	200
QB_2	1000	2000	0	1000	0	200
QB_3	0	1000	0	1000	−200	0
QB_4	1000	2000	0	1000	−200	0
OB_1	0	1000	0	500	0	200
OB_2	0	1000	500	1000	0	200
OB_3	1000	2000	0	500	0	200
OB_4	1000	2000	500	1000	0	200
OB_5	0	1000	0	500	−200	0
OB_6	0	1000	500	1000	−200	0
OB_7	1000	2000	0	500	−200	0
OB_8	1000	2000	500	1000	−200	0

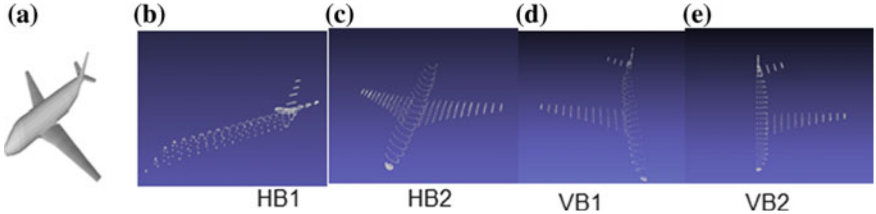


Fig. 3 **a** Aircraft model, **b** horizontal partitioned cloud HB1, **c** horizontal partitioned cloud HB2, **d** vertical partitioned cloud VB1, **e** vertical partitioned cloud VB2 (better viewed in color)

Beltami based shape extraction techniques for shape extraction and matching. These algorithms are a good basis for shape signature design.

We chose LBS because it is quite powerful due to its interesting properties such as invariant to scaling, rotation, and is also Isometric invariant with additional property as topology preserving capability. The principal direction of the surfaces of interest and geometrical information, such as localization of distension are provided by first and second Eigen function of Laplace Beltrami operator [15]. Isometric invariant holds since Eigen functions depend specifically on the gradient and divergence which are dependent on the Riemannian structure of the manifold. Similarly since Eigenfunctions are normalizable, scaling factor is also taken care of. Several variants of Beltrami based signatures are available in literature [13] processing, pose transfer, and parameterization. Another source of inspiration is where the eigenvalues of the same operator were used as a shape descriptor.

Let M be a normal manifold then Laplace Beltrami of M is given by [13]

$$\nabla_y f(y) = \mathrm{div}(\mathrm{grad}(f(y)))$$

The Laplacian can be calculated as the difference between the average of $f(y)$ and f on an infinitesimal sphere around f. For construction of shape descriptor several authors have used the Eigen functions of Laplace Beltrami. A general descriptor for this kind for Q-dimension is

$$f(y) = \sum_{i \geq 0} \tau(\lambda_i \varphi^2(y) \approx \sum_{i=0}^{l} \tau(\lambda_i)\varphi^2(y)$$

where $y_{i=0}^{I} \in M$, λ is eigenvalue of Laplace Beltrami, φ is Eigen function of Laplace Beltrami, $\tau(\lambda_i)$ is a transfer f function computed using Eigen values. These descriptors are re-computed for every vertex of point cloud and are computed efficiently by using I (small number) of Laplacian Eigen values and Eigen functions. It may be observed that the descriptor changes for every change of transfer function.

Heat Kernel Signature (HKS): This is a special case of LBS [13] where

$$\tau(\lambda) = \mathrm{e}^{-t_0 \lambda a^{\tau}}$$

$$\mathrm{HKS}(y) = \sum_{i=0}^{l} \mathrm{e}^{-t_0 \lambda a^{\tau_i}} \varphi^2(y)$$

and $\tau(\lambda)$ is a low-pass filter. The disadvantage of HKS by using a low-pass filter is its poor spatial localization.

Scale Invariant Heat Kernel Signature (SIHKS): As the major limitation of heat kernel signatures is its sensitivity to scale, in this work SIHKS [15] has been used as an alternate. For a given shape Y and its scaled version $Yo = \beta Y$, the new eigenvalues and Eigen functions will satisfy $\lambda_{\mathrm{o}} = \beta^2 \lambda$ and $\emptyset_{\mathrm{o}} = \beta \emptyset_{\mathrm{o}}$. If a shape is scaled β then the time shift will be $s = 2 \log_{\alpha} \beta$ and scaling of amplitude is given by β^2. In this proposed work, we remove β^2 by taking log of HKS and differentiate w.r.to τ then the resultant signature is computed as given below

$$\mathrm{hks}(y) = \frac{\sum_{i=0}^{I} -t_0 \alpha^{\tau_i} \log \alpha \mathrm{e}^{-t_0 \lambda \alpha^{\tau_i}} \varphi^2(y)}{\sum_{i=0}^{I} \mathrm{e}^{-t_0 \lambda \alpha^{\tau_i}} \varphi^2(y)}$$

$$\mathrm{SHKS}(\tau) = \mathrm{hks}(\tau + 1) + \mathrm{hks}(\tau)$$
$$\mathrm{SIHKS} = |\mathrm{FFT}(\mathrm{SHKS}(\tau))|$$

Wave Kernel Signature (WKS): This is also a particular case o LBS [13] given by

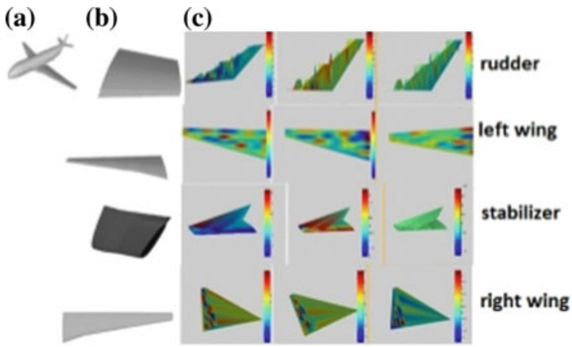


Fig. 4 Signatures of WKS for aircraft parts in the order of **a** first Eigen function, **b** second Eigen function and **c** fourteenth Eigen function

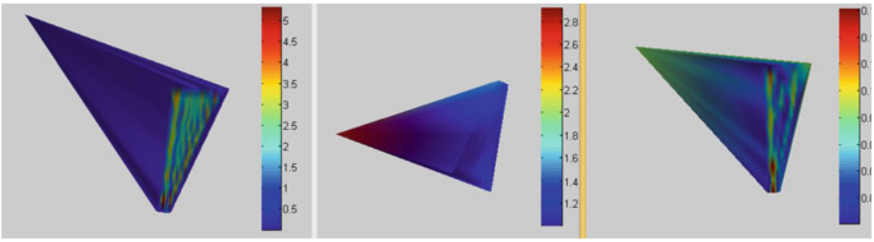


Fig. 5 Signatures for aircraft left wing for WKS, HKS and SIHKS respectively

$$\tau(\lambda) = \mathrm{e}\frac{(\log\ v_{1\,i} - \log\lambda)^2}{2\sigma^2}$$

$$\mathrm{WKS}(y) = \sum_{i=0}^{l} \mathrm{e}^{-\frac{\left(\log\ v_{1_i} - \log\ \lambda\right)^2}{2\sigma^2}} \varphi^2(y)$$

where v_1 is the mean energy level, $v_1 = \mathrm{func}(\lambda)$ and σ is the variance of kernel. WKS shape descriptor is based on the behavior of quantum particle on a manifold that a particle is having an initial energy distribution of log normal with mean energy level. The probability of finding an energy particle of v_1 at y is measured by WKS (Figs. 4 and 5).

C. *Deformation Estimation*

The signatures of the normal and the deformed cloud are different in size hence the missing values of the deformed signature can be computed using linear interpolation. The signatures are then compared using Euclidean distance and the visualization of deformation can be done based on different colors. The percentage of deformation is calculated if k the number of deformed vertices present in the deformed cloud is known prior.

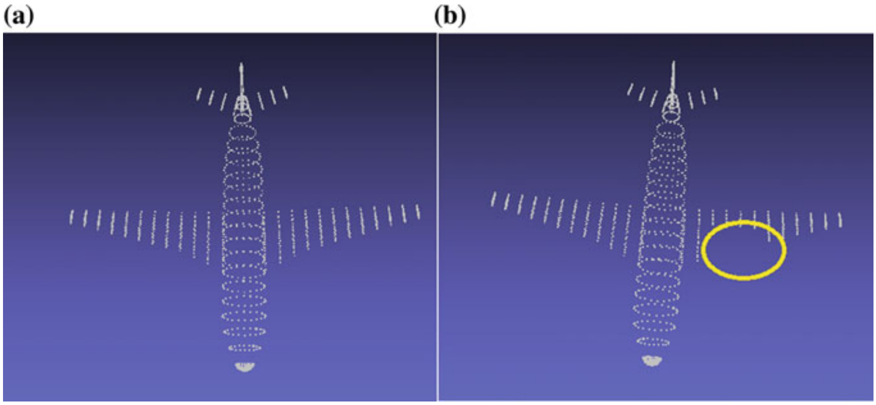
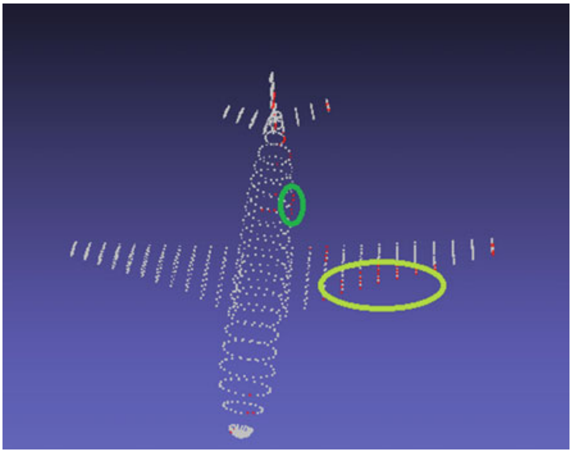


Fig. 6 Aircraft **a** normal, **b** deformed: Ground truth (better viewed in color)

Fig. 7 Deformation visualized in red color highlighted in yellow circle (better viewed in color)



The images of Fig. 6 the aircraft cloud without deformation and with deformation (marked in circle) in right wing is shown. The estimated deformation is shown with red color (marked in yellow circle) and false positive is marked in green circle is shown in Fig. 7.

D. *Segmentation*

One of the applications of LBS signatures is segmentation. Here the segmentation of right wing of the aircraft is shown in the red color by taking the first Eigen function of WKS. The median of first Eigen function is calculated and thresholding is applied to get the segmented cloud.

In Fig. 8 the segmentation is shown in red color. The visualization of the cloud with segmented color can be seen in part "a" of Fig. 8 but the segmented wing is better when the segmentation is done for partitioned cloud.

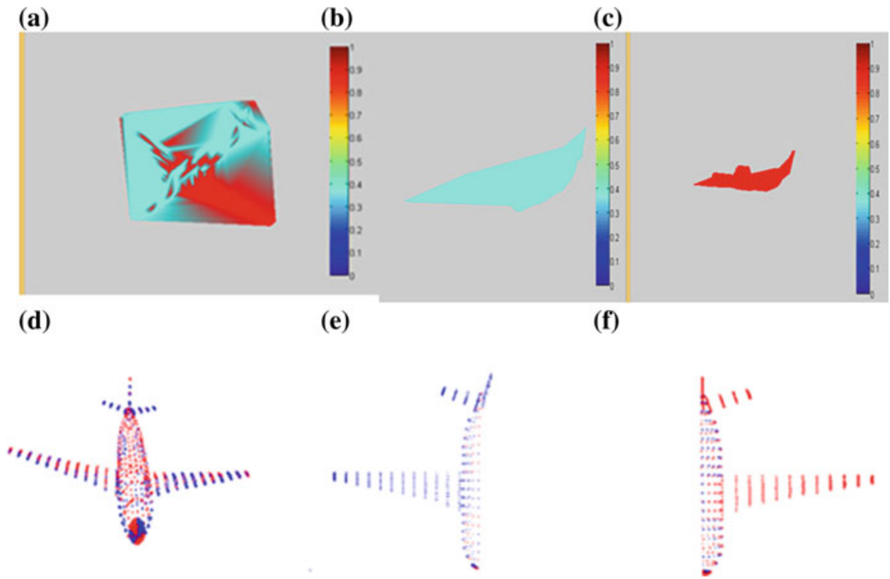


Fig. 8 **a** Wing segmentation (red color) of aircraft from WKS signatures, **b** OBB point cloud of the same segmentation of vertical bi–block 1, **c** the same segmentation of vertical bi–block 2 (**d**–**f**) point cloud representation in the same order

4 Conclusion and Future Work

In this work, we visualize the deformation of point cloud in red color along with segmentation. To get better estimation of deformation the interpolation can be bilinear and cubic. In future, this work can be extended by using various signatures and can match the clouds for deformation detection. In addition, this can be extended to calculate the correspondence between the clouds.

References

1. Andra S, Radke RJ, Roysam B, Al-Kofahi O (2005) Image change detection algorithms: a systematic survey. IEEE Trans Image Process 14(3):294–307
2. Wiselin Jiji G, Naveena Devi R (2015) Change detection techniques—a survey. Int J Comput Sci Appl (IJCSA) 5(2)
3. Konstantinos K (2014) Recent advances on 2D and 3D change detection in urban environments from remote sensing data. In: Computational approaches for urban environments, Geotechnologies and the environment. Springer, Cham, pp 237–272
4. Núñez P, Drews P, Bandera A, Rocha R, Campos M, Dias J (2010) Change detection in 3D environments based on gaussian mixture model and robust structural matching for autonomous robotic applications. In: IEEE/RSJ international conference on intelligent robots and systems (IROS). IEEE, pp 2633–2638
5. Vieira AW, Drews PL, Campos MF (2012) Efficient change detection in 3D environment for autonomous surveillance robots based on implicit volume. In: IEEE international conference on robotics and automation (ICRA). IEEE, pp 2999–3004
6. Monserrat O, Crosetto M (2008) Deformation measurement using terrestrial laser scanning data and least squares 3D surface matching. ISPRS J Photogram Remote Sens 63(1):142–154
7. Vaiapury K, Aksay A, Lin X, Izquierdo E, Papadopoulos C (2011) A vision based audit method and tool that compares a systems installation on a production aircraft to the original digital mock-up. SAE Int J Aerosp 4:880–892
8. Houda C, Thomas K, Pablo A, Peter R (2010) 3D change detection inside urban areas using different digital surface models. In: IAPRS, XXXVIII
9. Lu W, Vinutha K, Mayank B, Jayan E, Harpreet S, Denise P, Richard S (2014) Automatic 3D change detection for glaucoma diagnosis. In: WACV
10. Karthika R, Parameswaran L (2016) Study of Gabor wavelet for face recognition invariant to pose and orientation. In: Proceedings of the international conference on soft computing systems, Volume 397, 2016, of the series Advances in intelligent systems and computing, pp 501–509
11. Parameswaran L, Anbumani K (2008) Content-based watermarking for image authentication using independent component analysis. Informatica J 32:299–306
12. Chen L, Hoang D, Lin H, Nguyen T (2016) Innovative methodology for multi-view point cloud registration in robotic 3D object scanning and reconstruction. Appl Sci 6:132
13. Boscaini D, Masci J, Rodolà E, Bronstein MM, Cremers D (2016) Anisotropic diffusion descriptors. EUROGRAPHICS 35(2)
14. Levy B (2006) Laplace-Beltrami eigenfunctions towards an algorithm that "understands" geometry. In: Shape modeling and applications, international conference 2006, p 13
15. Kokkinos I, Bronstein MM (2010) Scale-invariant heat kernel signatures for non-rigid shape recognition. In: CVPR 2010
16. Arteaga RJ, Ruuth SJ (2015) Laplace-Beltrami spectra for shape comparison of surfaces in 3D using the closest point method. In: ICIP 2015
17. Buendia AE, Martinez LH, Vega RM, Torres JM, Crisóstomo ON (2015) Skeletons of 3D surfaces based on the Laplace-Beltrami operator eigen functions. Appl Math 6:414–420

A Study on Firefly Algorithm for Breast Cancer Classification

Harikumar Rajaguru and Sunil Kumar Prabhakar

Abstract One of the most serious and prominent cancers which affect more women in this world is breast cancer. Various kinds of breast cancer have been reported in medical literature that significantly differ in their capability to spread to other tissues in the human body. Plenty of risk factors have been traced through research though the exact reasons of breast cancer are not yet fully understood. Advances in medicine and technology help to improve the quality of life for breast cancer patients. A lot of contribution especially in the area of artificial intelligence and data mining is done to aid the diagnosis and classification of breast cancer. In order to provide assistance to the doctors, oncologists, and clinicians, neural networks and other optimization techniques serves as a great boon. Here, in this paper, firefly algorithm is utilized effectively as a powerful tool for the analysis and classification of breast cancer. Results show that an average classification accuracy of 98.52% is obtained when firefly algorithm is used for the classification of breast cancer.

1 Introduction

One of the cancers that develop from the tissue of the breast is known as breast cancer [1]. Breast cancer is a serious malignant tumor which arises from the cells of the breast. The significant signs and symptoms of breast cancer can include a red patch of skin, inverted nipples, oozing of fluid from nipples, skin dimpling, change in size and shape of the breast, and sometimes there could be a lump in the breast also [2]. If the breast cancer has spread, then the lymph nodes get swollen and there is a significant bone pain along with shortness of breath accompanied by fatigue [3]. The major factors and reasons for breast cancer in women are obesity and lack of nutritional diet. Lack of exercise, irregular menstruation, hormonal changes, and general lifestyle also contribute greatly to the occurrence of breast cancer. By the self-examination of the breasts using ultrasound testing, mammography, and biopsy, the diagnosis

H. Rajaguru (✉) · S. K. Prabhakar
Department of ECE, Bannari Amman Institute of Technology, Sathyamangalam, India
e-mail: harikumarrajaguru@gmail.com

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_42

of breast cancer can be easily done [4]. Depending on the type of cancer based on its position and severity and stages (0–4), the treatment options such as radiation, surgery, and chemotherapy can be decided easily by the doctor. Some of the common types of breast cancers include lobular carcinoma in situ, medullary carcinoma, adenoid cystic carcinoma, inflammatory breast cancer, mucinous carcinoma, invasive ductal, and lobular carcinoma, ductal carcinoma in situ, etc. [5]. With the aid of data mining, artificial intelligence, and soft computing, the classification procedure of breast cancer has become a boon for clinicians. A few significant and relevant works done by researchers for breast cancer classification are discussed as follows. For the diagnosis of breast cancer, feature selection combined with Support Vector Machine (SVM) was employed by Akay. The breast cancer classification was analyzed based on Radial Basis Functions (RBF) and Gaussian Mixture Model (GMM) by Rajaguru and Prabhakar [6]. A swarm intelligence-based wavelet neural network methodology for breast cancer detection and classification was reported by Dheeba et al. [7]. By the comparison of backpropogation training algorithm, the breast cancer can be classified easily by Paulin [8]. With the advantages of Linear Discriminant Analysis (LDA) and Soft Discriminant Classifier (SDC), the classification of breast cancer was done by Prabhakar and Rajaguru [9]. The multiclassifiers were utilized on various datasets for the diagnosis of breast cancer by Salama et al. [10]. For the breast cancer diagnosis and classification, a fuzzy-genetic approach was used by Pena-Reyes and Sipper [11]. Rajaguru and Prabhakar proposed an Expectation–Maximization Based Logistic Regression for Classification of Breast Cancer [12]. In this paper, firefly algorithm is used for breast cancer classification. The pictorial illustration of the paper is shown in Fig. 1.

2 Materials and Methods

As the first step, the clinical values obtained from the hospital are converted into numerical values. Chi-square test is implemented to all the numerical values. The values obtained by means of chi-square test are compared with different TNM stages of classification. The different stages of breast cancer are classified with the aid of firefly algorithm in this paper. For this study, a total of 82 patients are considered from the Department of Oncology of Kuppuswamy Naidu Memorial Hospital (KNMH), Coimbatore, Tamil Nadu, India. A lot of consolidation was done from hospital charts, referral letters, radiation therapy reports, radiological study reports, operative reports, and pathological reports. All the cancer patients were categorized into four stages based on their severity level. General parameters like food habits, personal hygiene, family history cancer, marital status, menstrual phase and duration, menopause, abortions (if any), size of breast, lump size and location, level and position of breast nipple, abdominal analysis, etc. were analyzed thoroughly. According to the standard International Union Against Cancer (UICC), 20 patients are present in the first stage and 23 patients are present in the final stage. Stage two has 28 patients and stage three has 11 patients.

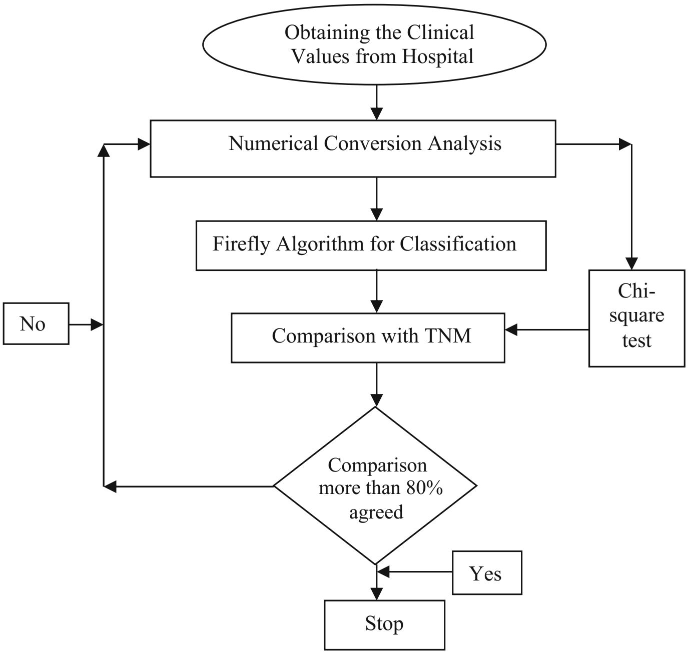


Fig. 1 Pictorial illustration of the work

2.1 Implementation of Chi-Square Test

A famous mathematical distribution which has a wide applicability in statistics oriented work is chi-square distribution [6]. To define this distribution, the Greek letter χ is used. The elements upon which this specific distribution is dependent are squared and so the symbol χ^2 is used to indicate the chi-square distribution. Every χ^2 distribution has an associated degree of freedom with it. The χ^2 statistic is pretty different from other statistics and has very little resemblance to the theoretical chi-square distribution. For the test of independence and for the goodness of fit test, the chi-square statistic is similar. For both the tests, all the categories into which the specific data is divided are utilized. The information obtained from the samples is the number of cases observed. These are known as the frequencies of occurrence for each and every category into which the information is grouped. In chi-square test, a statement is made by the null hypothesis dealing with the total number of cases to be expected in every category if this hypothesis is correct. For every category, the chi-square test is based on the difference between the observed and the expected values for all the categories.

The chi-square test is expressed as

Table 1 Performance measures of chi-square test for the 82 Breast Cancer Patients

	Performance of chi-square test		
	Minimum value	Maximum value	Average value
T1	3.84	19.52	11.7
T2	21.2	32.96	26.9
T3	30.72	40.85	35.5
T4	41.92	90.08	66

$$\chi^2 = \sum_i \frac{(Q_i - F_i)^2}{F_i}$$

where

Q_i denotes the observed number of cases in category i,
F_i denotes the expected number of cases in category i.

By the simple computation of difference calculation between the observed number of cases and the expected number of cases in every category, the chi-square statistic is used. This particular difference is squared and divided by the expected number of cases in a particular category. Then for all the different categories, these values are added and the total is referred to as chi-squared value. The null hypothesis bothers much about the distribution of data. For each chi-square test, the null and alternative hypothesis is stated as

$$H_0 : Q_i = F_i$$
$$H_1 : Q_i \neq F_i$$

If the null hypothesis claim is true, then the expected and observed values are pretty close to each other and $(Q_i - F_i)$ is pretty small for every category. Thus, when the null hypothesis claim is true, the chi-square statistic is small and when the null hypothesis claim is false, then the chi-square statistic is large. The degree of freedom is dependent on the total number of categories utilized in the calculation of the statistic. Thus, with the aid of chi-square distribution and chi-square statistic, one can determine easily whether the data is claimed or distributed. The performance analysis of chi-square test for the breast cancer patients is shown in Table 1.

3 Firefly Algorithm for Classification of Breast Cancer

The flashing phenomena of fireflies are simulated with the help of firefly algorithm and it was introduced by Yang [13]. In order to simulate the idea, the assumptions are made as follows. All the fireflies belong to the same gender. The fireflies which are brighter have a greater attraction than the fireflies which are less brighter. Between the fireflies if the distance increases, then there is a decrease in attraction. If none of the fireflies is brighter than a particular one, then the firefly will move randomly. The landscape of a particular objective function greatly influences the brightness of a firefly. For a specific maximization problem, the particular relation of the objective function and the brightness J is represented as $J(y)\alpha f(y)$. For the second assumption, the mathematical representation is as follows:

$$J(q) = (J_0)(\mathrm{e}^{-\gamma q^2})$$

where

J_0 denotes the actual density of light and
γ is named as the light absorption coefficient.
The attractiveness β is selected as follows:

$$\beta(q) = (\beta_0)\left(\mathrm{e}^{-\gamma q^2}\right)$$

where q denotes the distance between the two fireflies and β_0 represents the actual attractiveness at $q = 0$. The distance q_{jk} between the 2 fireflies at respective positions y_j and y_k is calculated as

$$q_{jk} = \left\| y_j - y_k \right\| = \sqrt{\sum_{m=1}^{d} \left(y_{j,m} - y_{k,m} \right)}$$

where $y_{j,m}$ is the mth component of the spatial coordinate y_j of the jth firefly. The specific movement of a firefly 'j' which is attracted to a much brighter firefly 'k' is determined as

$$y_j = y_j + \alpha\left(rand - \frac{1}{2} \right) + \beta_0 \mathrm{e}^{-\gamma q_{jk}^2}(y_k - y_j)$$

here, α denotes a random parameter in the interval [0, 1]. $Rand$ denotes a random number and is derived from a Gaussian distribution which is uniform in nature in [0, 1].

The Firefly Algorithm is given as follows:

(1) Objective function $f(y)$, $y = (y_1, \ldots, y_d)^T$
(2) A population of fireflies is initiated $y_j (j = 1, 2, \ldots, n)$
(3) The light intensity J_j at y_j is calculated by $f(y_j)$
(4) The light absorption coefficient γ is defined
(5) while ($t < Max\ Generation$)
(6) for $j = 1 : n$ all n fireflies
(7) for $k = 1 : n$ all n fireflies
(8) The distance $'r'$ is calculated between y_j and y_k using Cartesian distance Eq.

(9) if ($J_k > J_j$)
(10) Attractiveness varies with distance r through $\beta_0 e^{-\gamma q^2}$
(11) The firefly j is moved toward k in all the $'d'$ dimensions
(12) end if
(13) The new solutions are evaluated and the light intensity is updated
(14) end for k
(15) end for j
(16) The fireflies are ranked and the current best is found out
(17) end while

4 Results and Discussion

For the performance analysis of firefly algorithm as classifier for breast cancer classification , the standard benchmark parameters utilized for the analysis are sensitivity and specificity measures, classification accuracy, Perfect Classification Measures, False Alarm measures, Missed Classification Measures, and are given by the following mathematical formulae as:

$$\mathrm{PI} = \left(\frac{\mathrm{PC} - \mathrm{MC} - \mathrm{FA}}{\mathrm{PC}}\right) \times 100$$

PC is observed as Perfect Classification, MC is indicated as Missed Classification, and FA is denoted as False Alarm. The Sensitivity, Specificity, and Accuracy measures are mathematically expressed by the following formulae:

$$\text{Sensitivity} = \frac{\mathrm{PC}}{\mathrm{PC{+}FA}} \times 100$$

$$\text{Specificity} = \frac{\mathrm{PC}}{\mathrm{PC{+}MC}} \times 100$$

$$\text{Accuracy} = \frac{\text{Sensitivity+Specificity}}{2}$$

Table 2 depicts the performance analysis of firefly classifier for the various stages of breast cancer. Table 3 indicates the detailed analysis of comparison of clinical values with the firefly algorithm as classifier for the different stages of breast cancer classification.

5 Conclusion

Breast cancer is one of the most dangerous and the second leading reason of cancer death in women after lung cancer. In a most encouraging manner, the death rate from breast cancer has declined to a greater extent due to the advancement in medicine

Table 2 Performance analysis of firefly classifier for the different stages of breast cancer

	PC	MC	FA	PI	Sensitivity	Specificity	Average
T1	100	0	0	100	100	100	100
T2	100	0	0	100	100	100	100
T3	100	0	0	100	100	100	100
T4	88.23	11.76	0	86.67	100	88.23	94.11
Average	97.05	2.94	0	96.66	100	97.05	98.52

Table 3 Consolidate analysis of clinical values with firefly classifier for the different breast cancer classification stages

S.No	Breast cancer stage	Classification accuracy (%) through clinical procedure	Classification accuracy (%) through firefly algorithm as a classifier
1	T1	98	100
2	T2	100	100
3	T3	97	100
4	T4	100	94.11

and technology along with awareness. When cells in the tissue of the breast mutate and keep reproducing, then this problem occurs. These abnormal cells generally form a cluster and become a tumor. Some of the major risk factors of the breast cancer have been identified as age, gender, breast cancer gene mutation, change in size and shape of breast, race/ethnicity, hormonal fluctuations, weight, alcohol consumption, pregnancy history, radiation exposure, etc. In this paper, the firefly algorithm is utilized effectively as a classifier for breast cancer classification for all the four different stages and it is compared with the classification done through clinical procedure too. The classification accuracy of stages T1, T2, and T3 is 100% and for T4 stage, the classification accuracy is 94.11%. The average classification accuracy for all the stages from (0-4) produced is 98.52% and average performance index is 96.66%, along with an average specificity and sensitivity of 97.05% and 100%, respectively. Future works aim to make modifications in firefly algorithm for obtaining a better classification accuracy.

References

1. Polat K, Günes S (2007) Breast cancer diagnosis using least square support vector machine. Digit Signal Process 17:694–701
2. Setiono R (2000) Generating concise and accurate classification rules for breast cancer diagnosis. Artif Intell Med 18:205–219
3. Übeyli ED (2007) Implementing automated diagnostic systems for breast cancer detection. Expert Syst Appl 33:1054–1062

4. Mert A, Kılıç NZ, Bilgili E, Akan A (2015). Breast cancer detection with reduced feature set. Comput Math Methods Med 1–11
5. Rajaguru H, Prabhakar SK (2017) Bayesian linear discriminant analysis for breast cancer classification. In: 2nd IEEE international conference on communication and electronics systems. Coimbatore, India, pp 19–20 October 2017
6. Rajaguru H, Prabhakar SK (2016) A comprehensive analysis on breast cancer classification with radial basis function and gaussian mixture model In: 16th international conference on biomedical engineering (ICBME). Singapore, 7–10 December 2016
7. Dheeba J, Singh NA, Selvi ST (2014) Computer aided detection of breast cancer on mammograms: a swarm intelligence optimized wavelet neural network approach. J Biomed Inf 49:45–52
8. Paulin F (2011) Classification of breast cancer by comparing backpropagation training algorithm. Int J Comput Sci Eng 3:327–332
9. Rajaguru H, Prabhakar SK (2017) Performance analysis of breast cancer classification with softmax discriminant classifier and linear discriminant analysis. In: International conference on biomedical and health informatics. Thessaloniki, Greece, pp 18–21 November 2017
10. Salama GI, Abdelhalim MB, Zeid MA (2012) Breast cancer diagnosis on three different data sets using multi-classifiers. Int J Comput Inf Technol 1:36–43
11. Pena-Reyes CA, Sipper M (1999) A fuzzy-genetic approach to breast cancer diagnosis. Artif Intell Med 17:131–155
12. Rajaguru H, Prabhakar SK (2017) Expectation maximization based logistic regression for breast cancer classification In: IEEE proceedings of the international conference on electronics, communication and aerospace technology (ICECA 2017). Coimbatore, India, pp 603–606
13. Yang X-S (2009) Firefly algorithms for multimodal optimization. In: Proceedings of the international symposium on stochastic algorithms, vol 5792. Springer, Berlin, pp 169–178

Fuzzy C-Means Clustering and Gaussian Mixture Model for Epilepsy Classification from EEG

Harikumar Rajaguru and Sunil Kumar Prabhakar

Abstract Due to various disorders in the functionality of the brain, epileptic seizures occur and it affects the patient's mental, physical and emotional health to a great extent. The prediction of epileptic seizures before the beginning of the onset is pretty useful for seizure prevention by medication. One of the major causes for epilepsy is molecular mutation which results in irregular behaviour of neurons. Though the exact reasons for epilepsy are not known, early diagnosis is very useful for the treatment of epilepsy. Various computational techniques and machine learning algorithms are utilized to classify epilepsy from Electroencephalography (EEG) signals. In this paper, Fuzzy C-Means (FCM) Clustering algorithm is used as a clustering technique initially and then the features obtained through it is classified with the help of Gaussian Mixture Model (GMM) used as a post-classification technique. Results report that an average classification accuracy of 97.64% along with an average performance index of 95.01% is obtained successfully.

1 Introduction

Due to the unexpected and sudden surge of electrical activity in the brain, recurrent seizures occur [1]. The sudden electrical activities cause a huge distribution in the messaging system below the brain cells. Therefore, epilepsy is quite a serious and yet common neurological disorder with seizures as its primary symptoms [2]. Depending on the individual, the seizures can have a wide range of severity on the patients. Some common symptoms of epilepsy are: the patient falls down suddenly for no reason, occurrence of a convulsion with no fever, confused memory, inappropriate repetitive moments, sudden bouts of chewing and blinking, short spells of blackout, peculiar change in seizures such as smell and sound, rapid jerking movements and body stiffness [3]. Some of the similar symptoms often misdiagnosed as epilepsy include fainting, high fever, cataplexy, narcolepsy, panic attacks, nightmares, etc.

H. Rajaguru (✉) · S. K. Prabhakar
Department of ECE, Bannari Amman Institute of Technology, Sathyamangalam, India
e-mail: harikumarrajaguru@gmail.com

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_43

Epilepsy results when the messaging system in the brain is disrupted because of faulty electrical activity [4]. In many cases, the causes of occurrence of epilepsy are not known. Some patients developed epilepsy because of inherited genetic factors and epilepsy happens more often [5]. Some major factors that increase the risk of epilepsy is trauma in head, infectious diseases, serious brain conditions, prenatal injury, developmental disorders, etc. The various aspects of a person's life is affected terribly because of epilepsy which includes academic and work disturbances, unable to socialize well in the society and prohibits social development and interaction. EEG is the most famous physiological technique to record the electrical activity generated from the scalp of the brain [6]. Due to the synchronized activity of thousands of neurons, the electrical activity is generated and it is measured easily by EEG. EEG provides excellent time resolution and so it is widely used for the analysis of neurological disorders. A variety of works have been discussed in the literature regarding epilepsy classification from EEG signals. A clear understanding of the automated EEG analysis for epilepsy classification was done by Acharya et al. [7]. The epileptic seizure prediction in advance in a most efficient manner was done by Moghim and Corne [8]. With the aid of Adaboost Classifier and some important dimensionality reduction techniques, the epilepsy classification was performed with some interesting results by Prabhakar and Rajaguru [9]. Using Lyapunov exponent of the EEG signals, the characterization of epileptic seizures was done by Osowski et al. [10]. Based on computational intelligence techniques, a comparative study was made for epilepsy prediction and classification by Teixeira et al. [11]. The interpretation of Probabilistic Mixture Model (PMM) was discussed by Prabhakar and Rajaguru for the epilepsy classification from EEG signals [12]. The epileptic seizure prediction was performed with the relative spectral power features by Bandarabadi et al. [13]. Based on variational Bayesian GMM of zero-crossing intervals, the epileptic seizures were easily predicted in scalp EEG by Zandi et al. [14]. An efficient e-health system design for epilepsy classification was proposed by Prabhakar and Rajaguru [15]. In this work, Fuzzy C-Means Clustering and GMM algorithm were utilized well for epilepsy classification from EEG signals. The illustration of the work is shown in Fig. 1. The organization of the work is structured as follows. The materials and methods are explained in Sect. 2 followed by the usage of clustering technique in Sect. 3. Section 4 explains the classification using GMM algorithm followed by the results and discussion in Sect. 5 and ended with conclusion in Sect. 6.

2 Materials and Methods

The clinical EEG recordings of 20 epileptic patients who were admitted for their treatment from the Department of Neurology of Sri Ramakrishna Hospital, Coimbatore, India are taken for the analysis in this study [9]. The standard 10–20 International system was deployed to obtain the EEG recordings of the epileptic patients. A total of 16 channel electrodes were placed on the scalp of the epileptic patients and for various stages, the EEG recordings were easily measured. The recordings were measured

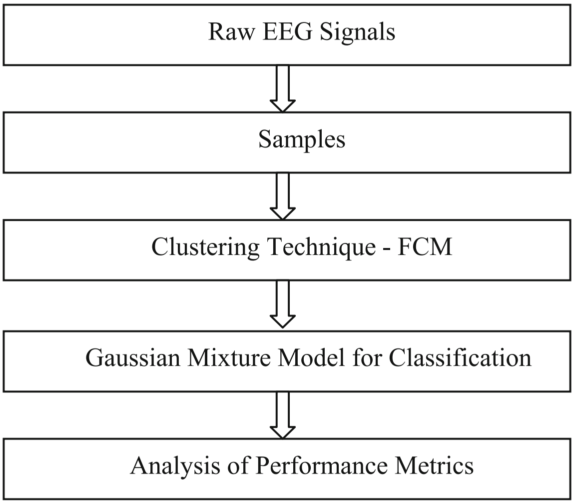


Fig. 1 Block diagram of the work

for a period of 55 min and it was obtained in European Data Format. Each channel was split into epochs and each epoch has the values pertaining to the instantaneous amplitude levels of the EEG signals.

3 Fuzzy C-Means Clustering Algorithm

To the raw EEG data and its samples, FCM algorithm is implemented to extract its features. Fuzzy Clustering algorithms have been of great utility to researchers and it is discussed and implemented in different areas [16]. FCM is one of the most commonly utilized algorithms. The total number of clusters used in this study 128. The dataset Z is divided into 'f' fuzzy clusters. This algorithm holds that every object belongs to a particular cluster but with a different membership degree. It implies that a cluster is considered as a fuzzy subset on the entire dataset.

Assuming that $Z = \{z_1, z_2, \ldots, z_n\}$ is a finite dataset, where $z_i = \left(z_{i1}, z_{i2}, \ldots., z_{iq}\right)$ denote a q dimensional object and z_{ij} is the jth property of the ith object. $F = \left\{F_1, F_2, \ldots, F_f\right\}$ indicates the f clusters. $W = \left\{w_1, w_2, \ldots, w_f\right\}$ represents $'f'$ l-dimensional cluster centroid points, where $w = \left(w_{i1}, w_{i2}, \ldots, w_{iq}\right).Y = \left(Y_{iy}\right)_{(n\times f)}$ is a fuzzy partition matrix and y_{ik} denotes the degree of membership of the i^{th} object in the gth cluster, where

$$\sum\nolimits_{g=1}^{f} y_{ig} = 1, \forall_i = 1, \ldots, n$$

The objective function is nothing but the quadratic sum of the weighed distance from all the samples to the cluster centroid in every cluster

$$P_r(Y, W) = \sum_{g=1}^{f} \sum_{i=1}^{n} y_{ig}^r d_{ig}^2$$

where $d_{ig} = \left\| z_i - w_g \right\|$ depicts the Euclidean distance between the ith object and the gth cluster centroid. $r[r \in (1, \infty)]$ denotes the fuzziness index which controls the fuzziness of the memberships. The membership becomes fuzzier when the value of $'r'$ becomes higher. FCM is carried out as an iterative process aiding in the minimization of the objective function P_r, with the updation of both Y and W. The steps are as mentioned below:

Step: 1 The initial value of the total number of clusters $'f'$, fuzziness index r, the threshold ξ and the maximum iteration $I_{\max}$ are assigned.
Step: 2 In a random manner, the fuzzy partition $Y^{(0)}$ is initialized based on the constraints of the membership degree.
Step: 3 The $'f'$ cluster centroids $W^{(t)}$ at the t-step is computed.
Step: 4 The objective function $P_r^{(t)}$ is calculated. If $\left| P_r^{(t)} - P_r^{(t-1)} \right| < \xi$ or $t > I_{\max}$, then stop otherwise continue to step 5
Step: 5 The $Y^{(t+1)}$ is calculated and returned to step 3. Finally, every object is arranged into one single cluster with accordance to the maximum degree of membership. The main advantages are that the algorithm is simple, has quick convergence capability and it is quite easy to extend.

The standard technique to determine the total number of clusters of FCM is as follows:

(1) The search range of the total number of clusters is set as the first step
(2) Run the FCM to generate the clustering results of various number of clusters
(3) Selection of a suitable clustering validity index
(4) Evaluate the clustering results
(5) Obtaining the optimal number of clusters based on the evaluation result.

The consolidated steps involved in the clustering procedure are as follows:

Step 1: Input the search range $\left[f_{\min}, f_{\max}\right]$
Step 2: For every integer $gn \in \left[f_{\min}, f_{\max}\right]$, run FCM and then the clustering validity index is calculated
Step 3: All the values of clustering validity index are compared. gn represents the optimal number of clusters f_{opt}
Step 4: The output is f_{opt}. It is the optimal value of clustering validity index and thus the clustering result is found out.

4 Gaussian Mixture Model as a Post Classifier

The cluster values are now fed inside the GMM classifier for epilepsy classification. Let $a \in \Re^d$ and G be the total number of components where every component has a prior probability d_i and probability density function with mean μ_i and covariance $\Sigma_i, i = 1, \ldots, G$.

A Gaussian Mixture Model is expressed as

$$\sum_{i=1}^{G} d_i \phi(a|\mu_i, \Sigma_i) = \sum_{i=1}^{G} d_i \frac{1}{\sqrt{(2\pi)^d |\Sigma_i|}} \exp\left(\frac{-(a-\mu_i)^t \Sigma_i^{-1} (a-\mu_i)}{2} \right)$$

where $\sum_{i=1}^{G} d_i = 1$

The likelihood function is expressed as

$$H(A|\theta) = \prod_{j=1}^{n} f\left(a_j|\theta\right)$$

and log-likelihood function is expressed as

$$l(A|\theta) = \sum_{j=1}^{n} \log\left(\sum_{i=1}^{G} d_i \phi\left(a_j|\mu_i, \Sigma_i\right)\right)$$

where $A = \left(a_1^t, \ldots, a_n^t\right)^t$, respectively.

Expectation Maximization was utilized for GMM [17]. The unobservable variable A in a particular space is indirectly observed through an observed variable Z in its respective sample space. Let $f(a|\theta)$ be the sampling density and it depends on the parameter $\theta \in \Omega$, therefore the corresponding family of sampling densities for Z is expressed as follows:

$$q(z|\theta) = \int_{a(z)} f(a|\theta) da$$

where $a(z)$ is a subset of A under the mapping $a \rightarrow z(a)$ from A to Z. The main objective of the EM algorithm is to find the value of θ so that $q(z|\theta)$ is minimized.

5 Results and Discussion

With FCM clustering and GMM algorithm, the parameters like Performance Index, Accuracy, Sensitivity and Specificity are computed in Table 1. The mathematical formulae for the parameters like Performance Index (PI), Sensitivity, Specificity and Accuracy are mathematically written as follows:

Table 1 Performance metrics analysis of FCM with GMM algorithm

Name of the parameter (%)	Average
PC	95.28
MC	0
FA	4.70
PI	95.01
Specificity	100
Sensitivity	95.28
Accuracy	97.64

$$PI = \left(\frac{PC - MC - FA}{PC}\right) \times 100$$

where

PC— Perfect Classification,
MC— Missed Classification and
FA— False Alarm. The Sensitivity, Specificity and Accuracy measures are expressed by the following equations:

$$Sensitivity = \frac{PC}{PC+FA} \times 100$$
$$Specificity = \frac{PC}{PC+MC} \times 100$$
$$Accuracy = \frac{Sensitivity+Specificity}{2}$$

6 Conclusion

Thus, epilepsy is a chronic neurological disorder causing recurrent and unprovoked seizures. If the seizure is stronger, then it causes uncontrollable muscle spasms and twitches and it lasts for a few minutes. The patients lose their consciousness or become quite confused if the seizure is stronger. EEG has a pivotal role to play in the diagnosis, classification and analysis of epilepsy classification. In this work, FCM is used as a clustering technique and later it is classified with the help of GMM. Results show that an average classification accuracy of about 97.64% is obtained, an average sensitivity rate of 95.28% is obtained, an average specificity rate of 100% is obtained and an average performance index of 95.01% is obtained. Future works aim to work with modified FCM and modified GMM algorithm for obtaining a better epilepsy classification rate.

Acknowledgements The authors are grateful to Dr. Asokan, Neurologist, Ramakrishna Hospital Coimbatore and Dr. B. Rajalakshmi, Diabetologist, Govt. Hospital Dindigul for providing the EEG signals.

References

1. Harikumar R, Kumar PS (2015) Fuzzy techniques and aggregation operators in classification of epilepsy risk levels for diabetic patients using eeg signals and cerebral blood flow. J Biomater Tissue Eng 5(4): 316–322
2. Rajaguru H, Prabhakar SK (2016) A framework for epilepsy classification using modified sparse representation classifiers and native bayesian classifier from EEG signals. J Med Imaging Health Inf
3. Prabhakar SK, Rajaguru H (2015) Analysis of centre tendency mode chaotic modeling for electroencephalography signals obtained from an epileptic patient. Adv Stud Theor Phys 9(4): 171–177, HIKARI Ltd., http://dx.doi.org/10.12988/astp.2015.5117
4. Harikumar R, Kumar PS (2015) Frequency behaviors of electroencephalography signals in epileptic patients from a wavelet thresholding perspective. Appl Math Sci 9(50): 2451–2457, HIKARI Ltd., http://dx.doi.org/10.12988/ams.2015.52135
5. Prabhakar SK, Rajaguru H (2016) Classification of epilepsy risk using variable thresholding based feature extraction technique and suitable post classifiers. Int J Simul Syst Sci Technol (IJSSST) 17(33): 28.1–28.8
6. Rajaguru H, Prabhakar SK (2017) Analysis of probabilistic neural networks with dimensionality reduction for epilepsy classification from EEG. Int J Mech Eng Technol
7. Acharya UR, Sree SV, Swapna G, Martis RJ, Suri JS (2013) Automated EEG analysis of epilepsy: a review. Knowl- Based Syst 45:147–165
8. Moghim N, Corne DW (2014) Predicting epileptic seizures in advance. PLoS ONE 9(6), Article ID e99334
9. Prabhakar SK, Rajaguru H (2017) Adaboost classifier with dimensionality reduction techniques for epilepsy classification from EEG. In: International conference on biomedical and health informatics. Thessaloniki, Greece, 18–21 November 2017
10. Osowski S, Swiderski B, Cichocki A, Rysz A (2007) Epileptic seizure characterization by Lyapunov exponent of EEG signal. COMPEL - Int J Comput Math Electr Electron Eng 26(5):1276–1287
11. Teixeira CA, Direito B, Bandarabadi M et al (2014) Epileptic seizure predictors based on computational intelligence techniques: a comparative study with 278 patients. Comput Methods Program Biomed 114(3):324–336
12. Prabhakar SK, Rajaguru H (2017) Conceptual analysis of epilepsy classification using probabilistic mixture models In: 5th IEEE winter international conference on brain-computer interface. South Korea, 9–11 January 2017
13. Bandarabadi M, Teixeira CA, Rasekhi J, Dourado A (2015) Epileptic seizure prediction using relative spectral power features. Clin Neurophysiol 126(2):237–248
14. Zandi AS, Tafreshi R, Javidan M, Dumont GA (2013) Predicting epileptic seizures in scalp EEG based on a variational bayesian gaussian mixture model of zero-crossing intervals. IEEE Trans Biomed Eng 60(5):1401–1413
15. Prabhakar SK, Rajaguru H (2016) Efficient wireless system for telemedicine application with reduced PAPR using QMF based PTS technique for epilepsy classification from EEG signals In: IFBME proceedings (Springer), international conference on advancements of medicine and health care through technology (MEDITECH), Romania, 12–15 October 2016
16. Saha I, Maulik U, Bandyopadhyay S (2009) A new differential evolution based fuzzy clustering for automatic cluster evolution. In: Proceedings of the IEEE international advance computing conference (IACC'09), IEEE, Patiala, India, 706–711 March 2009
17. Prabhakar SK, Rajaguru H (2016) Performance analysis of GMM classifier for classification of normal and abnormal segments in PPG signals In: 16th international conference on biomedical engineering (ICBME), Singapore, 7–10 December 2016

Analysis on Detection of Chronic Alcoholics from EEG Signal Segments—A Comparative Study Between Two Software Tools

Harikumar Rajaguru, Vigneshkumar Arunachalam and Sunil Kumar Prabhakar

Abstract Alcohol consumption is vulnerable to the brain and has a high risk of brain damage and other neurobehavioral deficits. This paper primarily focuses on massive data generated from EEG signals and its characterization with respect to various states of the human brain under influence of alcohol. A single trial 64-channel EEG database is utilized for classification of alcoholic states for a single patient. Singular Value Decomposition (SVD) features of EEG segments are computed. Even though EEG signals are acquired from alcoholic patient some of the EEG signal segments resemble EEG segments of normal, alcoholic, and epileptic persons. Depending on the SVD values, EEG segments are labeled as normal, alcoholic, and epileptic and then classified through Hard Thresholding and K-means clustering techniques. The classification is done using two different softwares in this paper, namely, MATLAB and R studio and then the results are compared. The results show that MATLAB software classifies better than R studio software with comparatively highest classification accuracy of 83.5% which is obtained when Hard Thresholding method is utilized.

1 Introduction

Alcoholism is one of the most common reasons for a lot of complications happening in the human body. Consuming alcohol for a very long time leads to larger health diseases for sure. The long-term use of alcohol damages all human organs and has a great toll on the human body [1]. Alcohol abuse causes blurred vision, impaired memory, and has a lot of brain deficits [2]. Electroencephalography (EEG) is an inter-

H. Rajaguru (✉) · S. K. Prabhakar
Department of ECE, Bannari Amman Institute of Technology, Sathyamangalam, India
e-mail: harikumarrajaguru@gmail.com

V. Arunachalam
Department of CSE, Bannari Amman Institute of Technology, Sathyamangalam, India

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_44

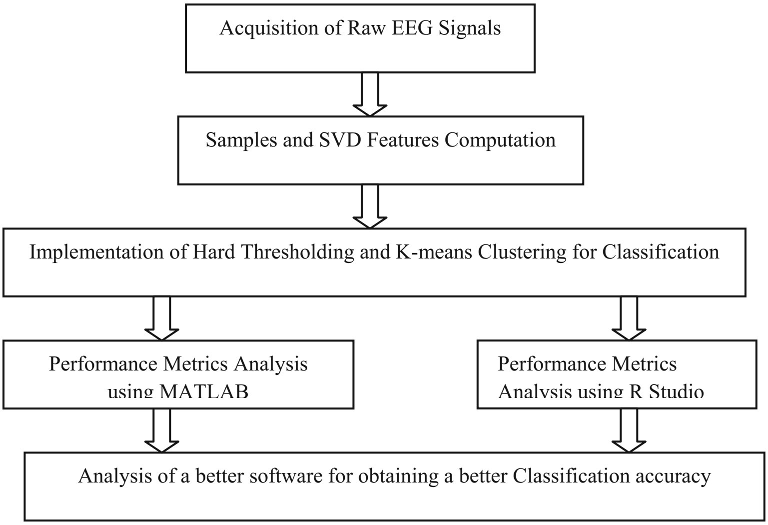


Fig. 1 Flow diagram for detection of chronic alcoholic segments from EEG signals through SVD labels using MATLAB and R studio

esting and vibrant electrophysiological monitoring technique utilized to record the electrical activity of the human brain [3]. EEG generally refers to the recording of the spontaneous electrical activity of the brain over a long/short specific period of time. In this work, the alcoholic EEG signal of a single patient is examined deeply with the help of two classification techniques and it is performed in two different softwares in order to analyze both the best software and the best classification technique.

Figure 1 depicts the flow diagram for the detection of chronic alcoholic segment from EEG signals through SVD labels. The alcoholic EEG data set is sampled and then mean, SVD features are calculated for each channels. These SVD values are used to label the channel through hard Thresholding and K-means clustering algorithms. The structure of the paper as follows: Sect. 1 introduces the manuscript. Section 2 discusses the materials and methods. Section 3 describes the classification procedure and the results are discussed in Sect. 4. Section 5 concludes the paper.

2 Materials and Methods

In this section, the acquisition of alcoholic EEG data acquisition and the feature extraction concept using SVD is discussed. Table 1 shows the details of the alcoholic EEG data [4].

Table 1 Details of the alcoholic EEG data

Total number of patients analyzed	1
Utilization of channel	Single trial 64 channel
Electrodes used	3
Signal sampling	256 samples/second
Analog to digital converter utilized	12 bit signed representation
Time required for every channel	10 s duration
Samples obtained per particular channel	2560
The total number of samples present for all the 64 signals	1,63,840
Total number of groups of samples	64

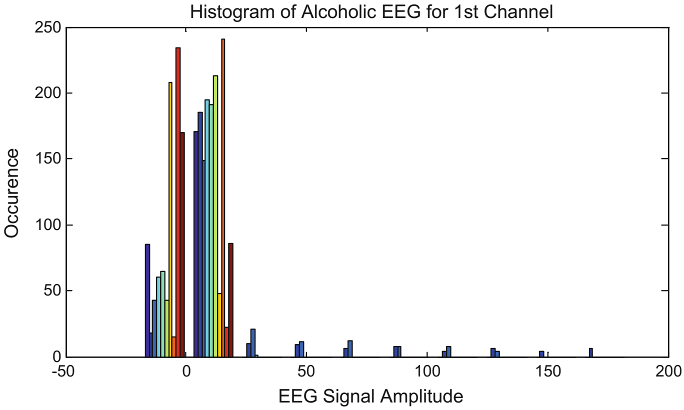


Fig. 2 Histogram of EEG signal amplitude for channel 1

2.1 Histogram Analysis

Histogram is a representation of repeated patterns based on their frequency of occurrence. It is observed from the analysis of amplitude components of EEG signals that nonlinear dynamics is responsible for this alcoholic state. The histogram of an alcoholic person for two different channels is depicted in Figs. 2 and 3. Figure 2 shows that the histograms for the first channel of alcoholic EEG signal which is skewed and non-Gaussian in nature, whereas the histogram of 64th channel of EEG signals as shown in Fig. 3 attained a nonlinear, non-Gaussian shape. Therefore, it is felt that the histogram may also be useful to differentiate the brain activity of an alcoholic patient on channel basis.

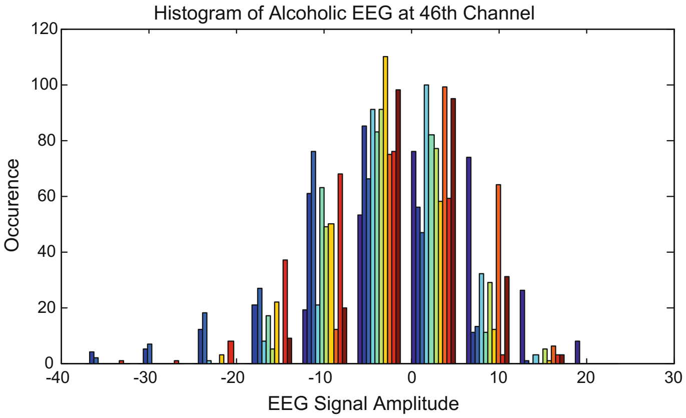


Fig. 3 Histogram of EEG signal amplitude for channel 64

2.2 Feature Extraction Using SVD

SVD plays a major role in most of the mathematical operations and is just a factorization of a real or complex matrix [5]. The SVD of a $l \times h$ matrix L is a factorization of the form $W \sum Y^*$, where W is a $l \times l$ unitary matrix (real or complex), $\sum$ is a $l \times h$ rectangular diagonal matrix and Y is a $h \times h$ unitary matrix (real or complex).

3 Classification Using Hard Thresholding and K-Means Clustering

The SVD features are classified with the help of Hard Thresholding classifier and K-means clustering Classifier to detect the risk of alcohol for the particular patient.

3.1 Hard Thresholding Classifier

Once the SVD features are computed, hard Thresholding classifier [6] is designed with the following conditions as mentioned in Table 2. The following representations are used.

μ_1 = the mean SVD of the entire normal segments of EEG signal
μ_2 = the mean SVD of the entire alcoholic segments of EEG signal
μ_3 = the mean SVD of the entire epileptic segments of EEG signal.

Based on the following constrains, the gold standard for the Thresholding classifier is arrived.

Table 2 Gold standard for thresholding classifier

Sl. no.	Mean SVD value of EEG segment	Clinical state
1	Up to μ_1	Normal
2	More than μ_1 and less than μ_3; with center at μ_2	Alcoholic
3	More than μ_2; with center at μ_3	Epileptic

Table 3 Classification through hard thresholding classifier for MATLAB and R studio

Platform classes	Alcoholic	Epileptic	Normal	Total
MATLAB	46	9	9	64
R studio	43	9	12	64

i. $\mu_1 < \mu_2 < \mu_3$;
ii. $(\mu_1 + \mu_2)/2 < \mu_2 < (\mu_2 + \mu_3)/2$;
iii. $(\mu_2 + \mu_3)/2 < \mu_3$.

Table 2 depicts the gold standard for Thresholding classifier based upon SVD values.

The algorithm is applied to a classifier and here they received pattern is a[n] = b[n] + d[n], where b(n) indicates an unknown pattern to be detected and d(n) indicates a white Gaussian noise. A lot of high-frequency features are present in the original pattern. The standard Thresholding techniques consist of both soft and hard Thresholding function.

Soft Thresholding is denoted as

$$Y(t) = \{a - \text{sgn}(a)T \text{ if } |a| \geq T;$$
$$0 \text{ if } |a| < T$$

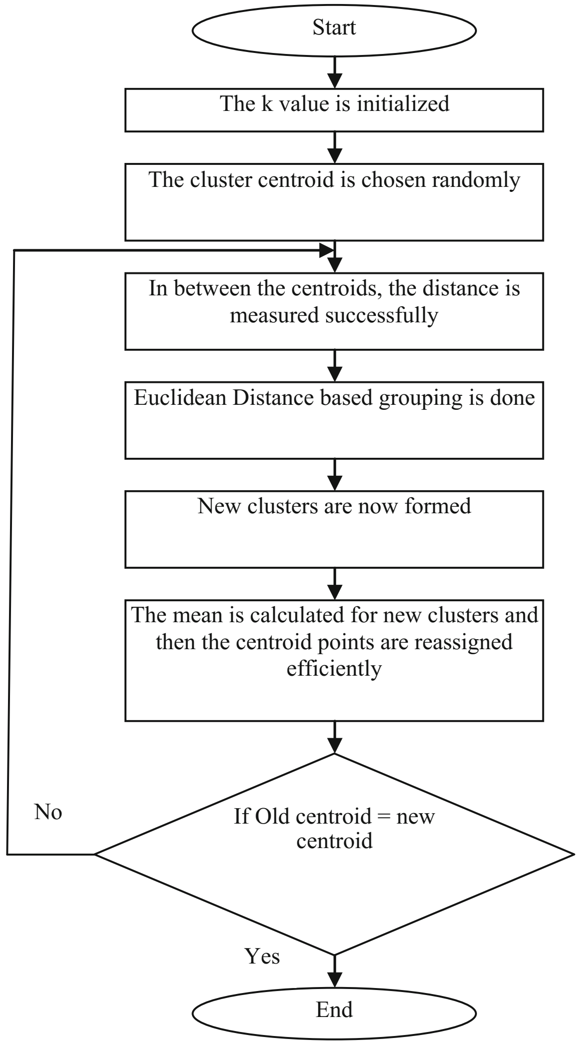
Hard thresholding is denoted as

$$Y(t) = \{ \ a \text{ if } |a| \geq T;$$
$$0 \text{ if } |a| < T$$

Table 3 shows the classification effectiveness of the alcoholic EEG data obtained through a single trial 64 channel using Hard Thresholding classifier for both MATLAB and R studio.

Out of the 64 channels, 46 channels were fully alcoholic, 9 channels were false classified as epileptic, and 9 channels were misclassified as normal by the MATLAB. For the R studio out of the 64 channels, 43 channels were fully alcoholic, 9 channels were false classified as epileptic, and 12 channels were misclassified as normal.

Fig. 4 Flow chart K-means clustering algorithm



3.2 K-Means Clustering

K-means clustering algorithm belongs to partitioning methods and is used widely because of its robustness, fast convergence, and simplicity [7]. Figure 4 shows the flow chart for K-means algorithm.

The basic process can be explained as follows:

First, the initialization of the K cluster centers is done and is chosen in a random manner.

Table 4 Classification through K-means clustering for the MATLAB and R studio

Platform classes	Alcoholic	Epileptic	Normal	Total
MATLAB	36	8	20	64
R studio	32	8	24	64

Second, each y_i is assigned to its nearest cluster center q_k by means of employing the Euclidean Distance (d).

$$KM(Y, Q) = \sum_{i=1}^{n} \min_{j \in \{1 \ldots k)} \left\| y_i - q_j \right\|^2$$

$$d(f, g) = d(g, f) = \sqrt{\sum_{i=1}^{n} (g_i - f_i)^2}$$

Third, each cluster center q_k is updated as the mean of all y_i that belongs to it.

Finally, unless the cluster centers have obtained stability, the steps are repeated.

Table 4 shows the classification usefulness of the alcoholic EEG data obtained through a single trial 64 channel using K-means clustering for both MATLAB and R studio.

As in the case of K-means clustering both the MATLAB and R studio classifies the data set in a similar manner. Out of the 64 channels, 36 channels were correctly classified as alcoholic, 8 channels were false classified as epileptic, and 20 channels were misclassified as normal in MATLAB software. Out of the 64 channels, 32 channels were correctly classified as alcoholic, 8 channels were false classified as epileptic, and 24 channels were misclassified as normal in R Studio software.

4 Results and Discussion

This paper identifies a good classifier and a better platform for an alcoholic EEG data set using SVD values and two different types of software. The performance of the classifiers as in this case of Hard Thresholding and K-means clustering are evaluated against their benchmark parameters like Performance Index, Accuracy, Specificity, and Sensitivity and the average results are computed in Table 5. The mathematical formulae for the Performance Index (PI), Sensitivity, Specificity, and Accuracy are given as follows:

$$\text{PI} = [\text{PC} - \text{MC} - \text{FA}/\text{PC}] * 100$$

Table 5 Average performance of SVD features with hard thresholding and K-means clustering in MATLAB software

Parameters (%)	Hard threshold as a classifier	K-means clustering as a classifier
Perfect Classification PC	71.8	40.62
Missed Classification MC	14.1	46.87
False Alarm FA	14.1	12.5
Performance Index PI	60.7	46.15
Specificity	83.5	46.43
Sensitivity	83.5	76.47
Accuracy	83.5	61.45

Table 6 Average performance of SVD features with hard thresholding and K-means clustering in R studio software

Parameters (%)	Hard threshold as a classifier	K-means clustering as a classifier
Perfect Classification PC	67.18	56.25
Missed Classification MC	14.1	31.25
False Alarm FA	18.72	12.5
Performance Index PI	51.15	55.47
Specificity	78.2	81.8
Sensitivity	82.65	64.28
Accuracy	80.42	73.05

where Perfect Classification is indicated by PC, Missed Classification is denoted by MC and False Alarm is specified as FA.

The Sensitivity, Specificity, and Accuracy measures are calculated as follows:

$$\text{Sensitivity} = [\text{PC}/(\text{PC} + \text{FA})] * 100$$
$$\text{Specificity} = [\text{PC}/(\text{PC} + \text{MC})] * 100$$
$$\text{Accuracy} = (\text{Sensitivity} + \text{Specificity})/2$$

Table 5 shows the classification worthiness of the both classifiers in the MATLAB platform. As tabulated in the Table 5, the Hard Thresholding method better classifies the data set with an accuracy of 83.5% when compared to the accuracy of 61.45% in K-means clustering method.

Average Performance of SVD features with Hard Thresholding and K-means clustering in R is shown in Table 6. It is observed from Table 6 that once again Hard Thresholding classifier scores better in the R studio platform with an accuracy of 80.42% when compared to the accuracy of 73.05% for K-means clustering algorithm.

5 Conclusion

Identification of valuable features and a better classifier is discussed in this paper for alcoholic EEG Signals of a single patient. SVD features are calculated and classified by using K-means clustering and Hard Thresholding. Softwares like MATLAB and R studio results are utilized and the results are compared. The results show that MATLAB software classifies better than R studio software with the comparatively highest classification accuracy of 83.5% is obtained when Hard Thresholding method is utilized. Further research will be in the usage of more heuristic and meta-heuristic classifiers for the analysis on detection of chronic alcoholics from EEG data segments.

References

1. Zou Y, Miao D, Wang D (2010) Research on sample entropy of alcoholic and normal people. Chin J Biomed Eng 29:939–942
2. Wu D, Chen ZH, Feng RF, Li GY, Luan T (2010) Study on human brain after consuming alcohol based on EEG signal. In: Proceedings of the 2010 3rd IEEE international conference on computer science and information technology, vol 5
3. Prabhakar SK, Rajaguru H (2016) Comparison of fuzzy output optimization with expectation maximization algorithm and its modification for epilepsy classification. In: International conference on cognition and recognition (ICCR 2016), Mysore, India, 30–31 December 2016
4. Prabhakar SK, Rajaguru H (2017) Softmax discriminant classifier for detection of risk levels in alcoholic EEG signals. In: IEEE Proceedings of the international conference on computing methodologies and communication (ICCMC July 2017), Erode, India
5. Prabhakar SK, Rajaguru H (2015) Performance comparison of fuzzy mutual information as dimensionality reduction techniques and SRC, SVD and approximate entropy as post classifiers for the classification of epilepsy risk levels from EEG signals. In: Proceedings of 2015 IEEE student symposium in biomedical engineering and sciences (ISSBES), 4 November 2015, Universiti Teknologi Mara, Malaysia
6. Zhang Q, Rossel RA, Choi P (2006) Denoising of gamma-ray signals by interval-dependent thresholds of wavelet analysis. Meas Sci Technol 731–735
7. Prabhakar SK, Rajaguru H (2015) PCA and K-means clustering for classification of epilepsy risk levels from EEG signals–a comparative study between them. In: Proceedings of the international conference on intelligent informatics and biomedical sciences (ICIIBMS), 28–30 November 2015, Okinawa, Japan

A System for Plant Disease Classification and Severity Estimation Using Machine Learning Techniques

Anakha Krishnakumar and Athi Narayanan

Abstract In India, more than 80% of agrarian crops are produced by smallholder farmers. The reports point that almost half the yield loss is mainly due to pests and diseases. Unlike pests, diseases are more difficult to detect and treat. Numerous studies and researches have been put forward to identify the behaviour of different diseases. Traditionally, farmers use naked eye observation for detecting disease but one of the areas considered today is processing the images with machine learning concepts to assist the farmers technologically. This paper presents an image processing strategy to classify sort of disease in a cucumber plant and gives a severity measure of malady spots in the cucumber leaf caught under real field condition.

1 Introduction

India is a developing country and around 70% of the people depend on agriculture. Agricultural profitability on production is something in which economy particularly depends. This is the one reason that disease detection in plants accepts an indispensable part in cultivation. The quantity, quality and productivity of crops are very important and hence proper care should be taken to resolve serious effects on plants.

The methods in image processing aim to recognize the leaf ailments. Otsu thresholding, K-means are the different segmentation techniques to separate the region of interest. Support Vector Machine (SVM), LDA, Neural System (NN) are the disease classification techniques to identify the affected symptoms on plants. The past works have a couple of demerits as lack of accuracy, lack of severity measure and lack of generic model [1].

A. Krishnakumar (✉) · A. Narayanan (✉)
Department of Computer Science and Engineering,
Amrita Vishwa Vidyapeetham, Amritapuri, India
e-mail: anakhakmr@gmail.com

A. Narayanan
e-mail: mail2athi@gmail.com

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_45

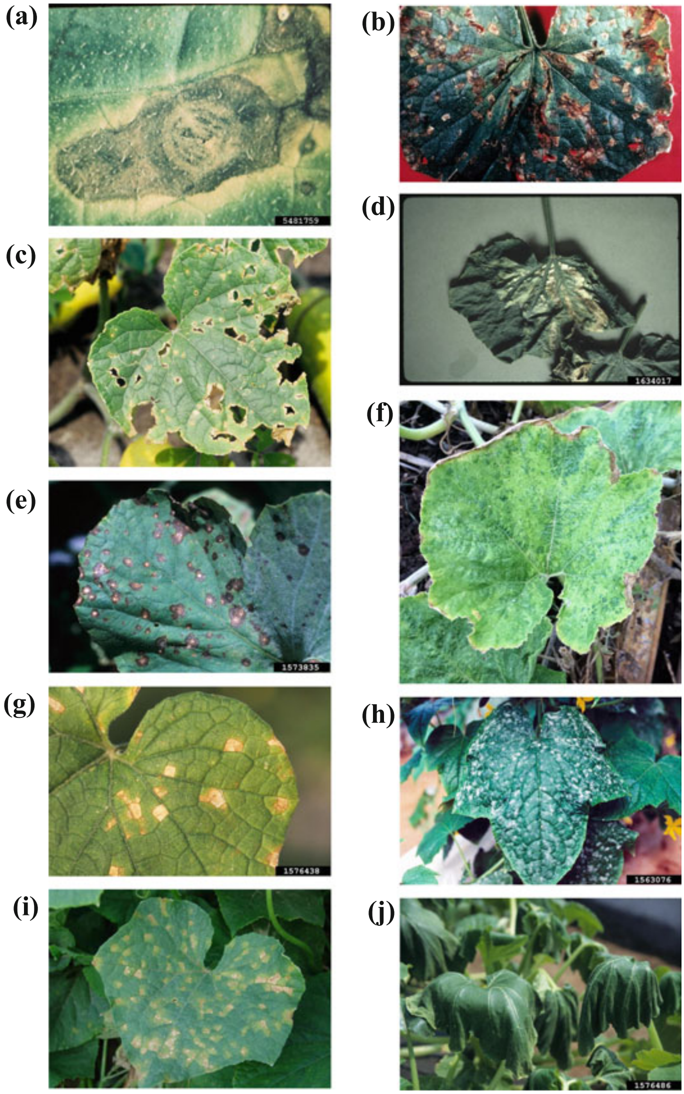


Fig. 1 Different types of affected leaf: **a** Alternaria leaf blight. **b** Angular leaf spot. **c** Bacterial leaf spot. **d** Bacterial Wilt. **e** Cercospora leaf spot. **f** Cucumber mosaic. **g** Target leaf spot. **h** Powdery mildew. **i** Downy mildew. **j** Phytophthora blight

Cucumis sativus is the scientific name of cucumber and is yearly creeping plant with sprawling vine, immense leaves and turning rings. Alternaria leaf blight, angular leaf spot, bacterial leaf spot, bacterial wilt, cercospora leaf spot, cucumber mosaic, target leaf spot, powdery mildew, downy mildew, phytophthora blight are the different kinds of diseases seen in cucumber leaf [2]. The following are mentioned in Fig. 1.

The paper is organized as follows: Sect. 2 specifies the related work. Section 3 introduces the proposed approach and Sect. 4 includes implementation and experimental results.
Section 5 concludes and specifies the future work.

2 Related Works

Lots of work in the field of detection of leaf diseases have been done by utilizing advanced image processing techniques. The methods include image acquisition, image preprocessing, image segmentation and feature extraction.

2.1 Image Processing

Initially, the technique of capturing the images in the RGB format are used for further processing and analysis purpose. The image taken under real field condition may contain inconsistency due to the lighting effects, which reduces the quality of picture. To deal with this situation, various preprocessing techniques are implemented such as mean filter, median filter and sharpening filter etc. To identify the distinctive portion in the image (the area of related pixels with equivalent properties), to discover the boundaries between the regions are the next important concerns. There are different strategies like K-means, Canny and Sobel division and ostu thresholding, etc. [3].

The preprocessing of image deals with removal of noise. There are different techniques to remove salt and pepper noise, Gaussian noise, Poisson noise, speckle noise, etc. [4]. The Fourier filtering transforms the digital image into frequency domain which implies the no. of frequency corresponding to no. of pixels in the spatial domain. Therefore, it is easy to examine the influence of particular pixels in an image and can be processed for further analysis. Sobel operator and Laplacian operator are used in edge detection. The edge detection aims to get the shape information of image where edge implies the sudden change in pixel values of an image [4]. The edge detection preprocessing which enhances the area of the image encloses edges and hence the sharpness of image increases, which makes the image clearer [5].

In the morphological preprocessing process the image is based on shape which mainly includes dilation and erosion operations. No of pixels added or removed from objects in the image depends on the size and shape of the structuring element [6]. Mainly concentrated to enable the selective enhancement of target structures. Gaussian filtering is used to blur the image to reduce the noise whereas in thresholding, it is used to classify the pixel values into different categories for later image analysis.

K-means is an unsupervised learning technique in which clusters with equivalent property are obtained. Initially, K-means picks k centroids and then reshape the clusters by assigning all data centre to the closest centroid of each cluster. In K-means we need to specify the number of clusters explicitly. But in otsu thresholding technique, all the possible threshold values are considered for choosing the one with minimum intraclass variance. The edges of diseased part are obtained by applying Canny and Sobel technique [7].

In order to get informative part from the segmented portion of the image, we predominantly choose shape, texture and colour features. These features are extracted mainly to identify the disease part from the image. Structural features give the information about size and shape of image whereas texture features deal with the intensity values [3].

2.2 Machine Learning

Machine learning algorithms are used to build the model which iteratively learns from the given information and forecast the future results. The area is broadly classified as supervised learning, unsupervised learning and reinforcement learning [7].

Support Vector Machine (SVM) classifies the data on the basis of decision planes modelled from the input samples. It can be linear classifier or nonlinear classifier depending upon the no of classes to be classified. In K-Nearest Neighbour classifier (K-NN), the training examples are classified in the view of nearest examples in the feature space [8]. Neural Network (NN) is another technique which uses the backpropagation system to classify the training examples but the system seems to be slower to train [7]. There are many kinds of classifiers used based on different features [9]. Effective results can be obtained from effective learning by training the model with large amount of dataset [10].

3 Proposed Method

3.1 System Design

The system mainly concentrates on early detection of plant disease. The basic diagrams for leaf disease detection are shown in Figs. 2, 3 and 4.

3.2 Disease Severity

Disease severity is the ratio of diseased area to total leaf area. The formula is given below:

$$\mathrm{DS} = \mathrm{DA}/\mathrm{LA} \tag{1}$$

where

DS Disease Severity

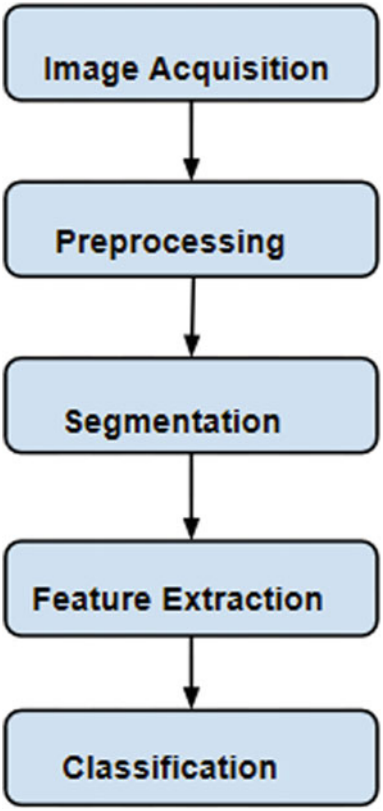


Fig. 2 System architecture

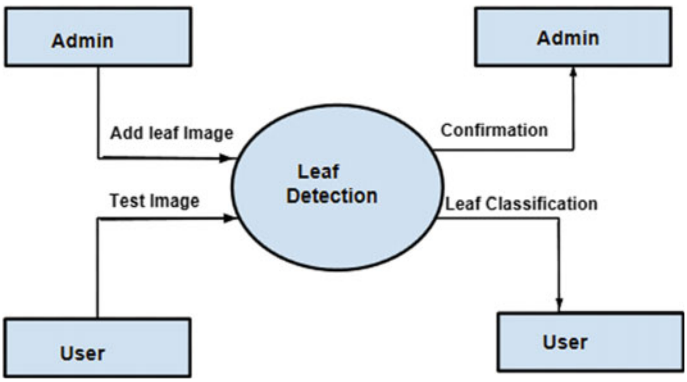


Fig. 3 Basic mechanism of the model

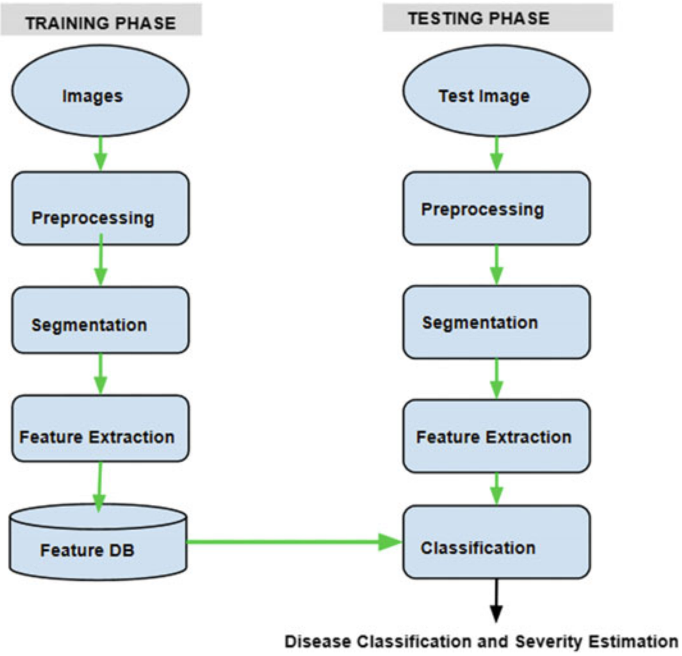


Fig. 4 Detailed workflow of model

DA No. of Pixel constitute disease area
LA No. of Pixel constitute normal leaf area.

3.3 Accuracy

The proportion of accurately recognized test images to the total number of test images and is given in the equation:

$$\%\text{Accuracy} = [CRT/TI] * 100 \quad (2)$$

CRT Correctly Recognized Test Images
TI Total Test Images.

4 Implementation and Experimental Results

Classification includes training and testing phase. The extracted feature values and its respective target values are utilized to train the model. Later this trained classifier

Fig. 5 Input and preprocessed image

model is used to predict the future results. Performance of SVM classifier is 86% when 50% of images from dataset is used for testing and the remaining for training.

4.1 Image Preprocessing

First step is the enhancing of the original image. Figure 5 shows original image and processed image.

Fig. 6 Segmentation using K-means clustering

4.2 Segmentation

The enhanced image is then segmented into three clusters. Figure 6 shows three clusters formed using K-means clustering.

4.3 Disease Classification and Disease Measure

The features include colour and texture which are extracted from the segmented images. In this work, ten features are figured for the sectioned parts of single leaf image [2]. These features, aggregately known as feature vector, and are utilized to train the SVM classifier. As a result, classifying the sample input leaf image into various classes. Features can be extracted using texture method or HOG to identify the different kind leaf disease and further using SVM for classification [9]. Figure 7 demonstrating the classification results.

Brightness refers to overall lightness or darkness of the image. Increasing the pixel value refers to lightening and decreasing the pixel value refers to darkening. Contrast is actually related to brightness where increasing the contrast makes light area lighter and dark area darker. Mathematically contrast refers to difference between maximum and minimum pixel. Images with high contrast are easier to recognize. As a part of feature analysis, it is important to understand how correlated a pixel to its neighbours. This can be measured based on the correlation factor [5].

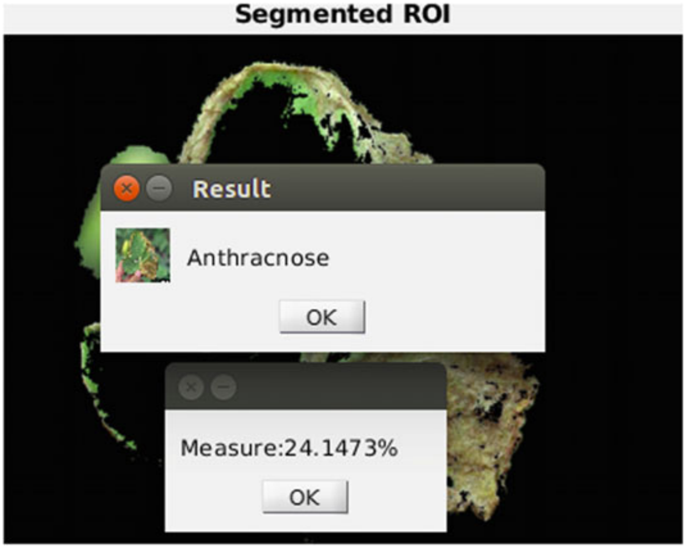


Fig. 7 Disease and its measure

We can analyze image based on pixel values since the image is composed of different ranges of pixels. Sometimes the values seem to be more similar or more random. To track these changes, there are different technical terms introduced. The measure of similarity of pixel refers to homogeneity and measure of randomness refers to the entropy. Uniformity of image refers to energy. Hence the homogeneity and energy are strongly correlated to each other. If the pixel seems to be more similar then the image will be more uniform and for a uniform image the value is set to 1. Therefore the image is a constant one and hence the contrast for constant image is 0.

Skewness is the measure of symmetry which implies the distribution on the left and right of the centre point is same. Negative value of skewness refers to skewed left implies left tail is long relative to right tail whereas positive value of skewness refers to skewed right and vice versa of skewed left. Kurtosis is the measure that deals whether the dataset is peaked or flat relative to normal distribution. The dataset with high kurtosis tends to have outliers whereas dataset with low kurtosis lack the outliers. The structural features provide information about the size and shape whereas texture features provides information based on the variation in intensities. The measure of image texture refers to IDM with 0.0 as highly textured and 1.0 as untextured [5] (Table 1).

Table 1 Image texture features

Features	Description
Contrast	Difference in brightness. Contrast is 0 for constant image
Correlation	How a pixel correlated to its neighbours. Range can be from −1 to +1
Energy	Measure of uniformity
Homogeneity	Measure of similarity of pixel
Entropy	Measure of randomness in pixel value
Variance	Indicates how each pixel varies from the mean pixel
Mean	Contribution of each pixel in the entire image
Standard deviation	Measure of deviation from mean
Skewness	Measure of symmetry where −ve value of skewness indicates left tail is long relative to right tail and vice versa
Smoothness	Measure of blurring to reduce noise
Kurtosis	If the dataset with high kurtosis tend to have outliers else lack of outliers
IDM	Measure of image texture where 0.0 is highly textured and 1.0 is untextured

Table 2 Management techniques

Disease	Cause	Pesticides
Alternaria leaf spot	Fungus	Chlorothalonil or mancozeb or copper fungicide
Anthracnose	Fungus	Chlorothalonil or mancozeb
Powdery mildew	Fungus	Sulphur or chlorothalonil or horticultural oil + baking soda
Downy mildew	Oomycete	Chlorothalonil or mancozeb or copper fungicide
Cercospora leaf spot	Fungus	Chlorothalonil or mancozeb
Bacterial wilt	Bacterium	no chemical control

5 Conclusion and Future Work

In this paper, we mainly focus on disease classification and severity measure which helps the farmers technologically to identify the leaf disease at the earliest. Also, we can identify the stage of affected plant based on the measure and classification and required management technique can be applied. This gives proficient and exact plant disease detection, discovery and characterization, measure figuring strategy by utilizing MATLAB. As a part of future expansion, we can bring out computerized model with the help of embedded system which can naturally splash the fertilizer by system spraying mechanism effectively at appropriate time (Table 2).

References

1. Sladojevic S, Arsenovic M, Anderla A, Culibrk D, Stefanovic D et al (2016) Deep neural networks based recognition of plant diseases by leaf image classification. Comput Intell Neurosci. http://dx.doi.org/10.1155/2016/3289801, Article ID 3289801, 11 pp
2. Prajapati BS, Dabhi VK, Prajapati HB et al (2016) A survey on detection and classification of cotton leaf diseases. In: International conference on electrical, electronics, and optimization techniques (ICEEOT)
3. Khirade SD, Patil AB (2015) Plant disease detection using image processing. In: 2015 International conference on computing communication control and automation
4. Thangavel SK, Murthi M et al (2017) A semi automated system for smart harvesting of tea leaves. In: 2017 International conference on advanced computing and communication systems (ICACCS-2017) 06, 07 Jan 2017, Coimbatore, India
5. Padol PB, Yadav AA et al (2016) SVM classifier based grape leaf disease detection. In: 2016 Conference on advances in signal processing (CASP), Cummins College of Engineering for Women, Pune. 9–11 Jun 2016
6. Singh V, Misra AK et al (2017) detection of plant leaf diseases using image segmentation and soft computing techniques. Inf Process Agric 4:41–49
7. Pujari JD, Yakkundimath R, Byadgi AS et al (2016) SVM and ANN based classification of plant disease using feature reduction technique. Int J Interact Multimed Artif Intell 3(7)
8. Mohanty SP, Hughes DP, Salath M et al (2016) Using deep learning for image-based plant disease detection, vol 7, September 2016. Article 1419
9. Venkataraman D, Mangayarkarasi N et al (2017) Support vector machine based classification of medical plants using leaf features. In: 2017 International conference on advances in computing, communications and informatics (ICACCI)
10. Anjali Rani KA, Supriya P, Sarath TV et al (2017) Computer vision based segregation of carrot and curry leaf plants with weed identification in carrot field. In: Proceedings of the IEEE 2017 international conference on computing methodologies and communication (ICCMC)

A Clinical Data Analytic Metric for Medical Ontology Using Semantic Similarity

Suraiya Parveen and Ranjit Biswas

Abstract Ontology is a set of concepts in a domain that shows their properties and the relations between them. Medical domain Ontology is widely used and very popular in e-healthcare, medical information systems, etc. The most significant benefit that Ontology may bring to healthcare systems is its ability to support the indispensable integration of knowledge and data (Pisanelli et al, Proceedings biological and medical data analysis, 6th international symposium, 2005, [1]). Graph structure is very important tool for Foundation, Analysis, and Domain Knowledge. Ontology as a graphical model envisages the process of any system and present appropriate analysis (Pedrinaci, Ontology-based metrics computation for business process analysis, [2]). In this study, the knowledge provided by the Ontology is further explored to obtain the related concepts. An algorithm to compute the related concepts of Ontology is also proposed in a simplified manner using Boolean Matrix. The inferences from this study may serve to improve the diagnosis process in the field of Biomedical Intelligence and Clinical Data Analysis.

1 Introduction

The Ontology is intended to accomplish a common and shared knowledge that can provide interoperability between computer applications, web pages, etc. ICT is becoming key foundation to improve healthcare by storing and analysing huge amount of data. ICT ensures its availability to all stakeholders in real time [3]. Medical ontologies are big support to realize the opportunity. Ontology is the most basic component of semantic technologies. This makes knowledge sharing fast and convenient. Interoperability, reuse and sharing ability of Ontology are powerful aspects

S. Parveen (✉) · R. Biswas
Department of Computer Science & Engineering,
SEST, Jamia Hamdard, New Delhi, India
e-mail: husainsuraiya@gmail.com

R. Biswas
e-mail: rbiswas@yahoo.com

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_46

to make semantic search possible [4]. The Ontology is visualized in the form of a graph structure. Graph forms the foundation for the design of many information processing systems such as transaction processing systems, decision support systems, project management systems, workflow systems, knowledge management, intelligent integration, search engine, and information retrieval [5]. The Ontology/schema is used to represent complex domain knowledge. Ontology represents knowledge of the domain with the help of "***IS-A***" relationship which is hierarchical [6]. In this study, Ontology is represented using Graph Structure in lieu Tree Structure. A graph represents Ontology more meritoriously in case of simple as well as complex domain. The nodes and link types are related through an Ontology graph [7].

> A Graph is defined by a pair G = {V, E} where V = {x1, x2, x3… xn} is a finite set of agile vertices and E a collection of edges that happen to connect these vertices.

Ontology is knowledge representation and to retrieve relevant information from the ontology is very essential. Similarity measurement in Ontology has been proposed by many researchers [8–12]. Boolean Matrix representation of graph structure is exploited to find the semantic similarity in Ontology. The idea is to convert Ontology in Boolean matrix form [6], and then use graph techniques to find semantic similarity of concepts. Ontology has sometimes been very complex; so it is useful to represent it in some simplified form.

In Sect. 1, we introduce the ontology and its representation. In Sect. 2, the approach of APO ontology, developed for K4Care project [13] is outlined by giving an overview. Section 3 elaborates the methodology of research and give the algorithm based on the study. In Sect. 4, we give an evaluation of our algorithm. Results are given in Sect. 5. Section 6 conclude the paper.

2 Case Study of APO Ontology

We are using Actor Profile Ontology (APO) to study the conceptual relationship in the Healthcare Domain. It has 399 concepts. APO Ontology is developed in the EU K4Care Project "Knowledge-Based Home care Service for an aging Europe". This project takes care of health issues of senior citizens in Europe. This is a knowledge-based e-health care project. The K4CARE [13] summarized below is a simple web-based model to provide home care services. In Fig. 1 the base concepts of APO Ontology are represented as hierarchal structure. These are the minimum elements required to provide basic home care. APO summarizes the contents of these concepts.

- Entity: An Entity is somebody who can perform an action on the K4Care project.
- Actor: The Actor class represents somebody in the K4Care. Actors are people interacting with the HC system. The class Actor is divided into subclasses such as Stable Members (Composed of Family Doctors, Physicians in Charge, Head Nurses, Social Workers, and Nurses), Additional Care Givers (Specialist Physician, Social Operator, Continuous Care Provider, and Informal Care Giver) and Patients.

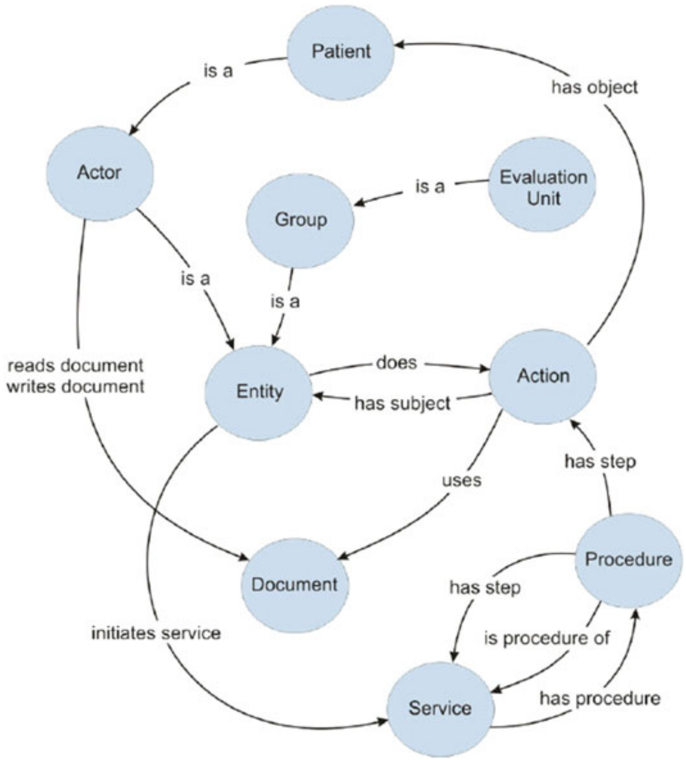


Fig. 1 APO base concept

- Evaluation Unit: This class represents a medical evaluation unit, a group that works together in the assessment and evaluation of patients. It is composed of a Family Doctor, a Physician Incharge, a Head Nurse, and a Social Worker.
- Services: The Service class represents the home care services. These services are classified into Access services, Patient Care services, and Information services.
- Procedure: This class represents a procedure associated with a service.
- Action: The class Action represents a set of simple actions that are executed inside one or more procedures.
- Document: This class represents the documents that are read or written during the execution of one action. These documents are stored in the EHR.

3 Methodology

In Ontology, we have considered only the taxonomical relation within a domain [6]. We require measuring semantic similarity within ontological concepts. Ontology is

Table 1 NCP matrix

	Patient	Actor	Entity	Action	Evaluation unit	Group	Document	Procedure	Service
Patient	#	1	0	1	0	0	0	0	0
Actor	1	#	1	0	0	0	1	0	0
Entity	0	1	#	1	0	1	0	0	1
Action	1	0	1	#	0	0	1	1	0
Evaluation unit	0	0	0	0	#	1	0	0	0
Group	0	0	1	0	1	#	0	0	0
Document	0	1	0	1	0	0	#	0	0
Procedure	0	0	0	1	0	0	0	#	1
Service	0	0	1	0	0	0	0	1	#

knowledge representation about a specific domain and presented in the form of the graph structure. Graphs are analytical tools which themselves give lots of information. In this paper, Ontology graph is manipulated with computation technique to find the sub-concept and super-concept of any particular concept (Ci). This is done for fast and simple retrieval of information from the Ontology.

This paper is an attempt to modify adjacency matrix representation of graph and using it for similarity measurement. Table 1 shows the Boolean matrix, which includes immediate successor and the immediate predecessor of the APO Ontology. This matrix leads to form a cluster of closely related concepts, which are interdependent on each other. The algorithm for the computation does not consider directions, but only adjacent pairs of a concept. We have named it **Neighbor Concepts Pair (NCP)**. To validate our results, we have used a statistical technique called Simple Matching Coefficient (SMC) [14].

3.1 Neighbor Concept Pair Matrix for APO Ontology

The aim of our work is to extract similar concepts within an Ontology. The above discussion about Ontology reveals that the predecessor and successor of any concept in the Ontology are related to each other being hierarchal. The following is the NCP matrix (n*n) of APO Ontology with nine concepts represented in Table 1. Each row contains 1 if the concept is adjacent (sub-concept/super-concept) with column concept, otherwise 0. Hash # symbol in the matrix shows same concept in row and column [15].

$$A_{ij} = \begin{cases} 1, & \text{If there is an edge from } v_i \text{ to } v_j \text{ or } v_j \text{ to } v_i \\ 0, & \text{otherwise.} \end{cases}$$

Table 2 Algorithm

Adjacency list

Input: Adjacency Matrix (m × n)

Output: Adjacent vertices of V_i

Step 1: Each Vertex maps to corresponding adjacent vertices (do not consider directions)

Step 2: Vertex List<adjacent vertices>

Step 3: To get all vertices adjacent to v_1,

Step 4: List<adjacent vertices> neighbors = adj.get(v_1)

// input value 1means i is adjacent to j

// input value 0 means i is not adjacent to j

Adjacency map (adjacency hash_table for concepts)

Each Vertex maps to a hashtable of adjacent vertices
Vertex (Vertex Edge)
To find out whether there's an edge from v_1 to v_2, call adj.get(v_1).containsKey(v_2)
To get the edge from v_1 to v_2, call adj.get(v_1).get(v_2)

3.2 Algorithm for Extracting Adjacent Clusters from NCP Matrix

An algorithm has been designed on above explanation of neighboring concept pair. This will extract sub-concepts and super concepts of any concept (C_i) in an Ontology. The algorithm is given in Table 2.

3.3 Clusters of Similar Concepts Generated by NCP Matrix

Ontology may be hierarchal or cyclic. In any case, the concepts in the Ontology are related to each other. When we represent the Ontology using proposed NCP matrix, it gives adjacent neighbors (subclass and super-class) of each concept in the Ontology, i.e., similarity cluster for each concept in Ontology. We obtain following related concepts, shown in Table 3 using this technique.

The semantically similar cluster has all nine basic concepts of APO ontology and corresponding row of each concept contain its related concepts.

4 Validation of Proposed Approach

Simple Matching Coefficient [14] is used for validation of our work. As per above matrix, APO Ontology shows the relationship between patient, actor, and action. Actor has relationship with patient, entity, and document. The cluster of similar concepts is shown in Table 3. To verify the similar concepts obtained from the Ontology using the proposed method; we have used Simple Matching Coefficient, as defined below:

Simple Matching Coefficient:

Simple Matching Coefficient (symmetric attributes) = number of matches/number of attributes

= (M 11 + M 00)/(M 01 + M 10 + M 11 + M 00)

The range of SMC is 0.00–1.00. The value SMC above 0.5 in a concept pair mean concepts are quite similar or related. SMC values below 0.5 means less similar concept pair.

We have applied the proposed method on APO Ontology and calculated Simple Matching Coefficient for concept pair as shown in Tables 4, 5, 6, and 7.

Table 3 Semantically similar cluster for APO ontology

Vertex	AV1	AV2	AV3	AV4	AV5
Patient	Actor	Action			
Actor	Patient	Entity	Document		
Entity	Actor	Group	Action	Service	Action
Action	Entity	Procedure			
Evaluation unit	Group				
Group	Evaluation unit	Entity			
Document	Actor	Action			
Procedure	Service	Action			
Service	Entity	Procedure			

Table 4 Similarity calculation patient and actor (neighboring concepts)

Patient	1	1	0	1	0	0	0	0	0
Actor	1	1	1	0	0	0	1	0	0

Table 5 Similarity calculation patient and action (neighboring concept)

Patient	1	1	0	1	0	0	0	0	0
Action	1	0	1	1	0	0	1	1	0

Table 6 Similarity calculation patient and document (neighboring concept)

Patient	1	1	0	1	0	0	0	0	0
Document	0	1	0	1	0	0	1	0	0

Table 7 Similarity calculation actor and evaluation unit (non-neighboring concept): SMC value for concept pair with no adjacency

Actor	1	1	1	0	0	0	1	0	0
Evaluation unit	0	0	0	0	1	1	0	0	0

5 Results

Simple Matching Coefficient is applied to NCP Matrix to verify the concept similarity within the Ontology. A minor modification is done so that the SMC can be applied. Hash # in NCP Table shows the same concept, but to find SMC it is taken as 1.

M 01 = 2 (the number of attributes where p was 0 and q was 1)
M 10 = 1 (the number of attributes where p was 1 and q was 0)
M 00 = 4 (the number of attributes where p was 0 and q was 0)
M 11 = 2 (the number of attributes where p was 1 and q was 1)

$$\begin{aligned} \text{SMC} &= (\text{M}\,11 + \text{M}\,00)/(\text{M}\,01 + \text{M}\,10 + \text{M}\,11 + \text{M}\,00) \\ &= (2 + 4)/(1 + 1 + 0 + 4) = 6/9 = 0.66 \end{aligned}$$

M 01 = 3 (the number of attributes where p was 0 and q was 1)
M 10 = 1 (the number of attributes where p was 1 and q was 0)
M 00 = 3 (the number of attributes where p was 0 and q was 0)
M 11 = 2 (the number of attributes where p was 1 and q was 1)

$$\begin{aligned} \text{SMC} &= (\text{M}\,11 + \text{M}\,00)/(\text{M}\,01 + \text{M}\,10 + \text{M}\,11 + \text{M}\,00) \\ &= (2 + 3)/(3 + 1 + 3 + 2) = 5/9 = 0.55 \end{aligned}$$

M 01 = 0 (the number of attributes where p was 0 and q was 1)
M 10 = 0 (the number of attributes where p was 1 and q was 0)
M 00 = 5 (the number of attributes where p was 0 and q was 0)
M 11 = 2 (the number of attributes where p was 1 and q was 1)

$$\begin{aligned} \text{SMC} &= (\text{M}\,11 + \text{M}\,00)/(\text{M}\,01 + \text{M}\,10 + \text{M}\,11 + \text{M}\,00) \\ &= (2 + 5)/(1 + 1 + 2 + 5) = 7/9 = 0.77 \end{aligned}$$

M 01 = 2 (the number of attributes where p was 0 and q was 1)
M 10 = 4 (the number of attributes where p was 1 and q was 0)
M 00 = 3 (the number of attributes where p was 0 and q was 0)
M 11 = 0 (the number of attributes where p was 1 and q was 1)

$$\begin{aligned} \text{SMC} &= (\text{M}\,11 + \text{M}\,00)/(\text{M}\,01 + \text{M}\,10 + \text{M}\,11 + \text{M}\,00) \\ &= (0 + 3)/(2 + 4 + 3 + 0) = 3/9 = 0.33 \end{aligned}$$

The similarity results for APO Ontology fulfil the four criteria of similarity measures [16]. These measures are as follows:

(1) Non-negativity: Similarity value cannot be less than zero.
(2) Identity: Sim (A, A) = Sim (B, B) = 1
(3) Symmetry: If any attribute of A is similar to attribute of B, Then same attribute of B is similar to same attributes of A.
Sim (A, B) = Sim (B, A)
(4) Uniqueness: Sim (A, B) = 1 → A = B

The range of semantic similarity is from zero to one.

The above similarity result clearly shows that neighboring concept pair have high similarity values. Hence, they are more similar as compared to non-neighboring concept pair.

6 Conclusion

ICT is becoming a very valuable asset for healthcare, Patient's Medical Records, Drug Records, Schedules, Diagnosis, Test Report, and many related issues are maintained and retrieved in real time. The accuracy, timely reminders, fast processes, decision support tool for diagnosis are most important in Healthcare; the proposed technique of finding semantically similar clusters in this work can be very useful for the same. It is crucial to avoid any human errors, missing information, related disease to help patients under the observation of many doctors. The work can be used as a decision support tool in healthcare.

References

1. Pisanelli DM, Pinciroli F, Masseroli M (2005) The Ontological lens: zooming in and out from genomic to clinical level. In: Proceedings biological and medical data analysis, 6th international symposium, Nov 2005
2. Pedrinaci D (2009) Ontology-based metrics computation for business process analysis, Knowledge Media Institute, The Open University

3. Riañoa D, Joan FR, López A, Sara VF, Patrizia E, Roberta M (2012) Annicchiaricod carlo caltagironede: an ontology-based personalization of health-care knowledge to support clinical decisions for chronically ill patients. J Biomed Inform 45(3):429–446
4. Pérez G, López F, Corcho O (2004) Ontological engineering, 2nd printing. Springer, Berlin. ISBN: 1-85233-551-3
5. Biswas R, Gaur D (2008) Fuzzy meta node fuzzy metagraph and its cluster analysis. J Comput Sci 4(11):922–927. ISSN: 1549-3636
6. Batet M,Valls A, Giber K (2008) Measuring similarities in ontology by means of Boolean metrices, Intelligent Technologies for Advanced Knowledge Acquisition, Department of Computer Engineering and Maths, Universitat Rovira i Virgili
7. Barthélemy M, Chow E, Eliassi-Rad T (2005) Knowledge representation issues in semantic graphs for relationship detection, U.S. Department of Energy by University of California Lawrence Livermore National Laboratory, UCRL-CONF-209845
8. Wu Z, Palmer M (1994) Verb semantics and lexical selection. In: 32nd annual meeting of the association for computational linguistics, pp 133–138
9. Leacock C, Chodorow M (1998) Combining local context and WordNet similarity for word sense identification. In: Fellbaum C (ed) WordNet: an electronic lexical database. MIT Press, Cambridge, pp 265–283
10. Al-Mubaid H, Nguyen HA (2006) New ontology-based semantic similarity measure for the biomedical domain. In: Proceedings of the IEEE conference on granular computing, GrC-2006, Atlanta, GA, 10–12 May 2006, pp 623–628
11. Pedersen T, Patwardhan S, Michelizzi J (2004) WordNet: similarity –measuring the relatedness of concepts. AAAI, pp 1024–1025
12. Zhang X, Jing L, Hu X, Ng M, Zhou X (2007) A comparative study of ontology based term similarity measures on PubMed document clustering. In: Advances in databases: concepts, systems and applications DASFAA, vol 4443, pp 115–126
13. K4CARE, www.k4care.net
14. Segaran T (2007) Programming collective intelligence: building smart Web 2.0 applications. Beijing, O'Reilly
15. Samantha D (2001) Classic data structure. PHI
16. Veltkamp RC, Latecki LJ (2006) Properties and performances of shape similarity measures

True Color Image Compression and Decompression Using Fusion of Three-Level Discrete Wavelet Transform—Discrete Cosine Transforms and Arithmetic Coding Technique

Trupti Baraskar and Vijay R. Mankar

Abstract In this research paper, we have done the implementation and analysis of true color image compression and decompression technique. The implemented paper divides the color image into RGB component then after applying three-level Discrete Wavelet Transform, RGB components are split into nine higher frequency sub-bands and one lower order sub-band. The lower frequency sub-band is compressed into T-Matrix using One Dimension Discrete Cosine Transform. At the same time, higher frequency sub-bands are compressed using scalar quantize and eliminate zero and store data algorithm are applied to remove zeros in sub-band matrixes. Last, the encoded mode adopted arithmetic encoding. This algorithm has use two level of quantization this show significance improve in performance of compression algorithm. The decompression process is reverse process of encoder. The decompression algorithm decoded high-frequency subbands using return zero matrix algorithm and recover low-frequency sub-bands and other sub-bands using applying inverse process.

1 Introduction

Now a days, the use of image in decompression algorithm. The visual lossless compression is a practice that was presented by the imaging industry and it is trying to tell their users that encoding error are not noticed and viewing by persons. Maximum computer graphic adapter cards practice a true color model for display. It uses 8 bits per channel. Most graphics file formats accumulate 24-bit image where three

T. Baraskar (✉)
Department of Electronic Engineering, SGBAU, Amravati, India
e-mail: trupti.baraskar@mitpune.edu.in; baraskartn@gmail.com

V. R. Mankar
Department of Electronics Engineering, Government Polytechnic, Amravati, India
e-mail: vr_mankar@rediffmail.com

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_47

component red, blue, and green has 8 bits each. This produces probable 16 million colors. A true color array can be of less unit8, unit16, single or double. In a true color array, each color component is a value between 0 and 1. If pixel has color values are (0, 0, 0) is displayed as black and if color value are (1, 1, 1) is displayed as white. A pixel can represent three color components in three dimensions of the data array. An example is RGB (5, 2, 1).

1.1 Image Format Used to Store RGB Image

The visual information that can be stored in digital computer has two main classes. The vector class presents the image information that can be described by simple geometric function, i.e., segment, color, data bands, circumferences. The raster class deal with the basic element of image, i.e., pixel value. The raster class arranges the pixel value in a pre-defined order that is later used by compression schemes.

Generally, raster class contains a representation of a graphic stored as pixels at a fixed position with a fixed resolution. The popular available raster formats are GIF, JPEG, PNG, TIFF, and BMP. The three dimension voxel raster class are used in medical imaging such as MRI Scanners. Another commonly used format is PBM (Portable Bitmap). This classifies into three different image format; binary, grayscale, and color images. This format has a common structure with uncompressed size [1, 2].

The proposed work is organized as follows. First in Sect. 2, we describe the existing techniques for true color image compression and decompression. In Sect. 3, there is a description of block diagram for compression and decompression, Sect. 4 has display result with a brief discussion. Finally, Sect. 5 is dedicated to the conclusion of implemented work.

2 Related Work

The compression of the true color image has various surveyed image compression approaches. The true color image compression algorithms are categorized into lossy or lossless compression schemes. The outmoded lossless compression method; RLE (Run-Length encoding) method has given better compression measure on the specific data with a large number of redundant information. In the true color image, the data has continuous pixel for one color line and this continuity can break down for multiline of the same color. Thus, the traditional RLE cannot guarantee to get reliable compression measures [3]. Thus, a new approach combined the DCT transform with improving RLE coding is implemented to achieve better compression ratio within acceptable range of data loss [4]. The Discrete Wavelet Transform (DWT) gives a high degree of correlation between RGB planes of the true color image. By combining Genetic Algorithm (GA) with DWT, more suitable space planes are generated. This result is found to be superior in terms of quality of the reconstructed image [5].

The available lossless compression methods include Shannon–Fano coding, Huffman coding, LZW (Lempel Ziv Welch) coding, and arithmetic coding. The Lossy compression methods contains predictive coding, block truncation, transform, vector quantization, sub-band, and fractal coding. Out of all abovementioned methods, discrete cosine transform is often used in the realization of transform coding for the image pixel [6–8].

The literature survey is based on transform approaches used by available compression techniques like discrete cosine transform and discrete wavelet transform. A new lossy compression algorithm has two transforms (DCT and DWT) with T-matrix coding and optimal sub-band threshold by using SURE and neigh shrink. The method enhances compression performance of image without reducing the PSNR quality [9, 10]. The served hybrid scheme combing the DWT and DCT transform which gives high compression ratio. In this paper, simulation work shows constant improved performance as compared to the JPEG-based DCT and the Daubechies-based DWT [11].

3 Proposed Work

In this paper, we have used fusion of transforms for implementation of color image compression. Mostly, color image has integer value of the pixels. Thus, for the compression, a luminance–chrominance model is considered where luminance represents the intensity of the image and looks like a grayscale and the chrominance components represent color information in the image [12, 13]. The color components of an 8-bit RGB image are integers in the range between 0 and 255 rather than floating-point values in the range [0, 1]. The proposed method works on luminance–chrominance models using DWT–DCT Transform. In this proposed method, image format used to store RGB image is Portable Pix Map [14].

3.1 Block Diagram of Compression

This compression approach is split into two parts; first, a three-level DWT is applied to a color image (A), which splits the image into nine high-frequency sub-bands (HL3, LH3, HH3, HL2, LH2, HH2, HL1, LH1, HH1) and one low-frequency sub-band (LL3). The high-frequency sub-bands at level one, two and three are quantized and encoded directly by EZSD. Second, the low-frequency sub-band are compressed by DCT and quantized then encoded by arithmetic coding (Fig. 1).

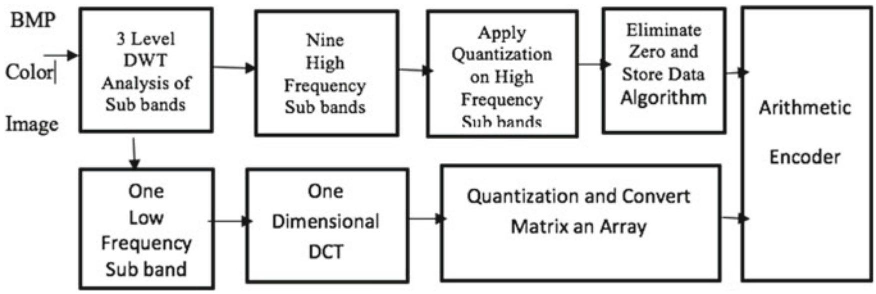


Fig. 1 The block diagram of compression

3.1.1 Three-Level DWT Decomposition

If the 2D analysis filter bank with Haar wavelet filter is applied on the digital image, the result is a first-level decomposition of Discrete Wavelet Transformed (DWT). The first level of decomposition has four parts, i.e., LL1, LH1, HL1l and HH1 sub-bands. It indicates the approximated image and the images with horizontal, vertical, and diagonal boundaries separately. The LL1 sub-band further to have a second and third level of DWT decomposition. The pixels within each sub-band represents the DWT coefficients corresponding to the sub-band. Mostly, the LL sub-band at the three level of decomposition has significant information as compared to all other sub-bands The approximated image has high-scale, low-frequency components and the detail of image has low-scale, high-frequency components. These results determine excellent energy compaction properties of DWT. This property can archive image compression. Normally, we have to transmit the image in limited bandwidth for compression and hence the DWT coefficients are to be quantized and efficiently encoded (Fig. 2).

3.1.2 Apply Quantization on Higher Frequency Sub-Bands

Quantization in an image compression is an example of lossy technique. Quantization is achieved by many-to-one mapping technique that replaces set of value into one value. There are two basic methods of quantization; SQ (scalar quantization) performs many-to-one mapping and VQ (vector quantization) replaces each block of input pixels with the index of a vector in the codebook. The decoder simply receives each index and looks up the corresponding vector in the codebook [15, 16]. In this proposed compression algorithm, we utilized high-frequency sub-bands as input vector to quantize using Scalar Quantization. Now Scalar Quantization performs many-to-one mapping on each value. The true value image contains standard quality factor and it varying [0.01–0.1] [17].

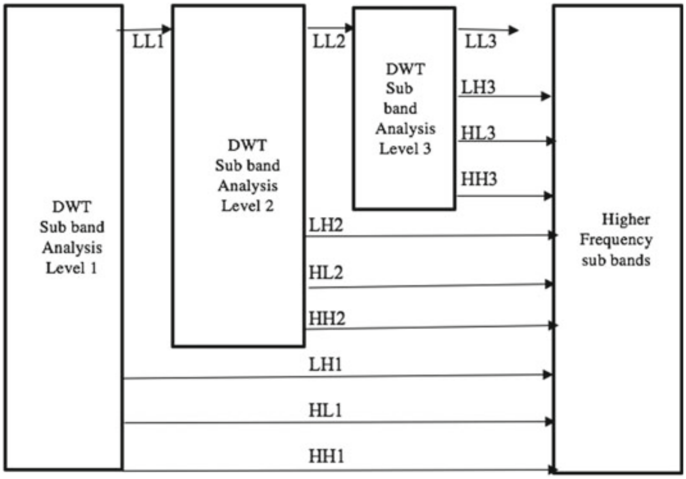


Fig. 2 The decomposition process of three-level DWT

Table 1 Evaluation of quantized level based on different quality value (Qv) and standard quality factor (Q)

Quantization level	If quality value (Qv) = standard quality factor (Q)	If quality value (Qv) > standard quality factor (Q)	If quality value (Qv) < standard quality factor (Q)
$Qm = Q \cdot 2^{m-1}$	$Qm = Qv \cdot 2^{m-1}$ if $Qv = 0.01$ and $Q = 0.01$	$Qm = Qv \cdot 2^{m-1}$ if $Qv = 0.05$ and $Q = 0.01$; $Qv = 0.1$ and $Q = 0.01$; $Qv = 0.2$ and $Q = 0.01$;	$Qm = Qv \cdot 2^{m-1}$ if $Qv = 0.005$ and $Q = 0.01$
Q1 = Q	0.01	0.05, 0.1,0.2	0.005
Q2 = 2Q	0.02	0.1, 0.2, 0.4	0.01
Q3 = 4Q	0.04	0.2, 0.4, 0.8	0.02
Q4 = 8Q	0.08	0.4, 0.8, 1.6	0.04
Q5 = 16Q	0.16	0.8, 1.6, 3.2	0.08

If quantizes Value > 0.01 or 1% then quantized level of high-frequency sub-band are as following:

Table 1 shows the evaluation of the quantization level for different relation between Quality Value (Qv) = 0.01, 0.05, 0.1, 0.2, 0.05 where Standard Quality Factor = 0.01. Table 2 indicates that the output parameter of Quantization level depends on Quality Value. If Qv $\geq$ Q; Image Quality low but we can achieve high compression ratio and vice versa (Fig. 3).

Table 2 Comparison of output parameter for selection of quality value (Qv)

S. No.	Condition for selection of quality value (Qv)	Output parameters	
		Image quality	Compression ratio
1	If quality value (Qv) > standard quality factor (Q)	Low	High
2	If quality value (Qv) < standard quality factor (Q)	High	Low
3	If quality value (Qv) = standard quality factor (Q)	Accept level	Accept level

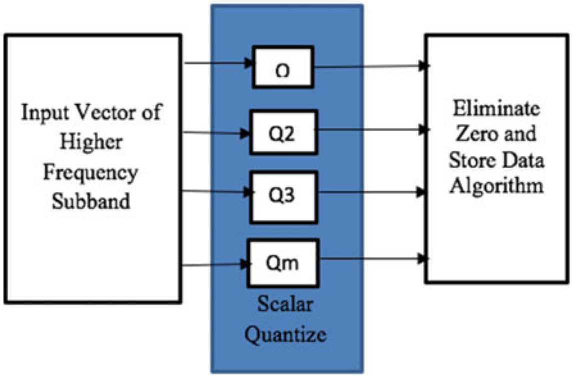


Fig. 3 The quantization process on higher frequency sub-bands

3.1.3 Apply Eliminate Zero and Store Data Algorithm on High-Frequency Sub-Bands

As level decomposition increases, sub-bands undergo partition and information details shift toward left corner (LL bands), meanwhile high-frequency sub-bands become more insignificant. Hence, the ignorance of this insignificant information is not perfect solution toward image quality. If quantized (If Quality Value (Qv) >= Standard Quality Factor [Q = 0.01]), high-frequency sub-bands have a rich number of zeroes, unnecessary coding of these zeroes make the algorithm complex and time-consuming. Therefore, EZSD is used to eliminate blocks of these zeroes and to store the blocks of nonzero data [15]. The EZSD algorithm is useful to increase compression ratio for high-frequency sub-bands. This algorithm is mainly applicable for eliminating zero and store any one nonzero data in a reduced array (Fig. 4).

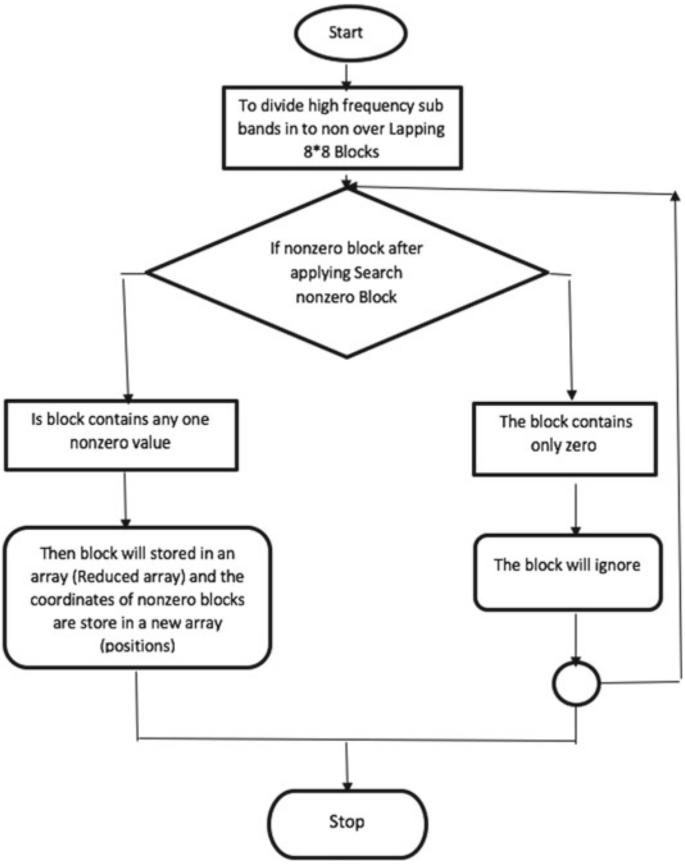


Fig. 4 The flow chart of eliminate zero and store data algorithm on high-frequency sub-bands

3.1.4 LL3 Sub Band Compression Using One-Dimensional DCT, Quantization Converted to an Array

The one-dimensional DCT and quantization process is applied to every row and column of LL3 sub-band array. The discrete cosine transforms (DCT) helps to separate the image into spectral sub-bands. Each sub-band has different importance (with respect to the visual quality of image). The equation for one dimension (N data items) DCT is as following (Fig. 5):

The equation of One-Dimensional DCT

$$X_c(k) = (1/N) \sum_{n=0}^{N-1} X_n \cos(k 2\pi n/N) \quad \text{where}\, k = 0, 1, 2 \ldots N-1$$

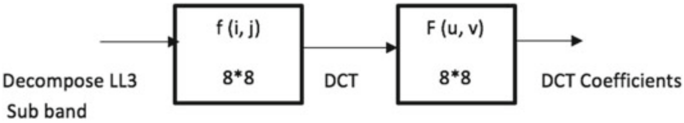


Fig. 5 One-dimensional DCT

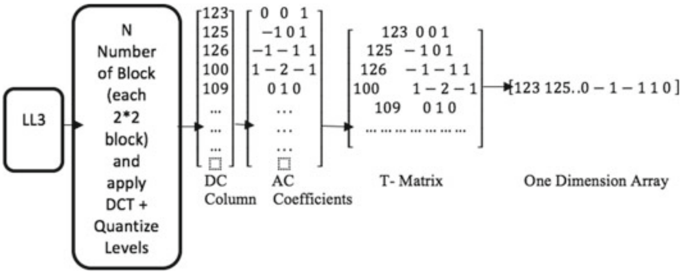


Fig. 6 The flow of implementation for quantizing levels and T-Matrix

The equation of One-Dimensional IDCT

$$X_c(k) = \sum_{n=0}^{N-1} C(u) X_n \cos(k2\pi n/N) \quad \text{where } k = 0, 1, 2 \ldots N-1$$

$$C(u) = 1 \quad \text{for } u = 0 \text{ and } C(u) = 2 \quad \text{for } u = 1, 2, 3, \ldots N-1$$

The LL3 sub-band coefficients are difficult to encode directly by arithmetic coding because of its properties of DCT-like De-correlation Property, Energy Compaction, and integer value. The proposed image compression algorithm using DCT and arithmetic encoding for true color image model are as follows: The LL3 sub-band is divided into 2 * 2 blocks of matrix.

1. Apply DCT on each 2 * 2 blocks of matrix.
2. The DCT produce DC and AC coefficients. Every DC coefficient is stored in the DC column matrix and, respectively, other three AC coefficients are stored in AC matrix.
3. The DC column is partitioned into 64 or 128 parts, then transformed value and stored in the matrix as row, called Transform Matrix (T-Matrix).
4. Then, scan T-Matrix column by column and convert them into one-dimensional array.
5. Apply compression process using arithmetic coding (Fig. 6).

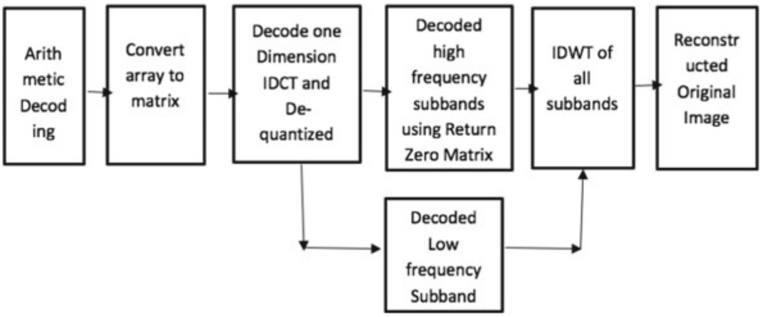


Fig. 7 Block diagram of decompression

3.2 Block Diagram of Decompression

See Fig. 7.

4 Result Analysis

The result shows that lower the MSE higher the PSNR and lower the SNR and better the compression ratio, out of five image4.bmp gives good result (Table 3).

Table 4 shows that if Quality value (Qv) > Quality Factor (Q) we get Higher PSNR and better compression ratio. Out of four third case gives good results (Figs. 8, 9, 10 and 11).

Table 3 Our compression and decompression methods result for five image inputs

Image name	Original image size (KB)	Compressed image size (KB)	Decompressed image size (KB)	PSNR	MSE	SNR	Compression ratio
image1.bmp	1352	79	Decompression not working	NA	NA	NA	17.11392405
image2.bmp	732	47	732	31.5364	59.182	26.4046	15.57446809
image3.bmp	769	38	732	35.5833	58.4021	29.7125	20.23684211
image4.bmp	2026	103	2026	30.1916	51.0853	22.3542	19.66990291
image5.bmp	769	35	769	35.992	60.5467	31.5013	21.97142857

Table 4 Comparison of image2.bmp, compression ratio for various quantization levels

	Original image size (KB)	Compressed image size (KB)	Decompressed image size (KB)	PSNR	MSE	SNR	Compression ratio
If quality Qv = Q; Qv = 0.01 and Q = 0.01	732	79	732	32.0296	59.182	26.8979	9.265822785
If quality Qv > Q; Qv = 0.05 and Q = 0.01	732	56	732	31.891	59.182	26.7592	13.07142857
If quality Qv > Q; Qv = 0.1 and Q = 0.01	732	47	732	31.5364	59.182	26.4046	15.57446809
If quality Qv < Q; Qv = 0.005 and Q = 0.01	732	63	Decompression not working	NA	NA	NA	11.61904762

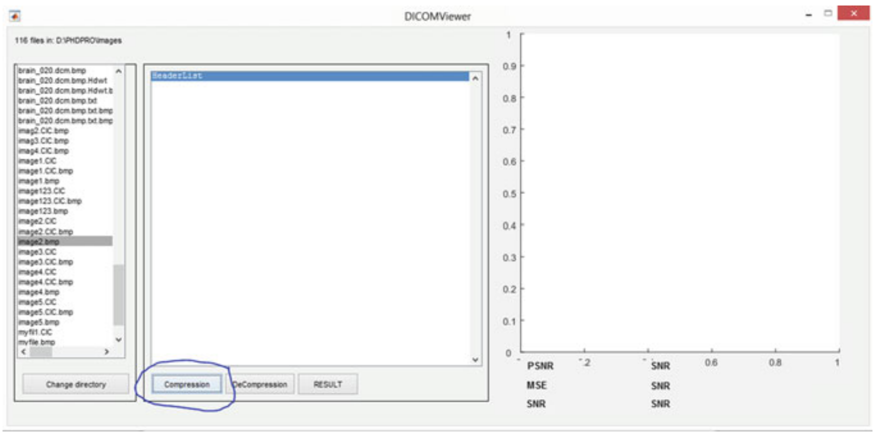


Fig. 8 Graphical user interface of DICOM viewer

5 Conclusion

This paper introduces new compression and decompression technique using fusion of three-level Discrete Wavelet Transform and Discrete Cosine Transforms with T-matrix coding. This implement helped to increases number of high-frequency coefficients and also to rises compression ratio. It also shows a method which is useful for quantization-level selection for different condition. This approach identifies if the quality value is greater than the standard quality factor, we can achieve higher compression and maintain image quality. The implemented method increase number

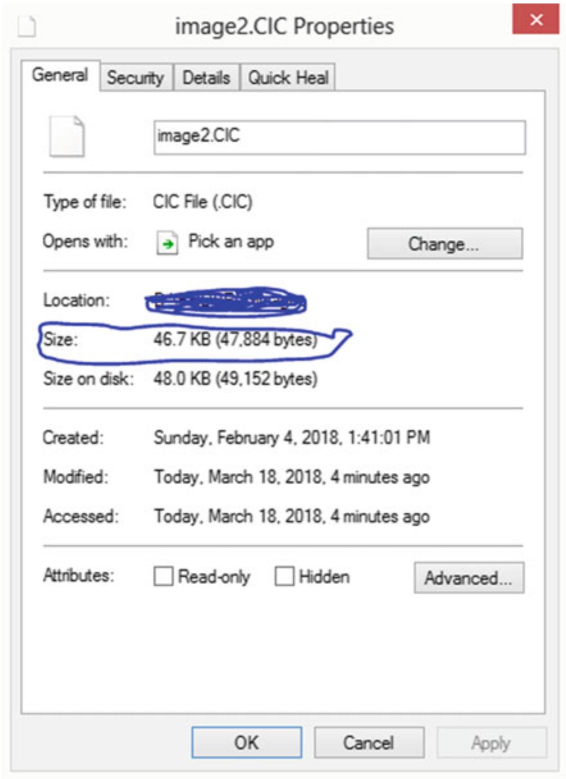


Fig. 9 Graphical user interface of image2.CIC with compression size 47.884 bytes

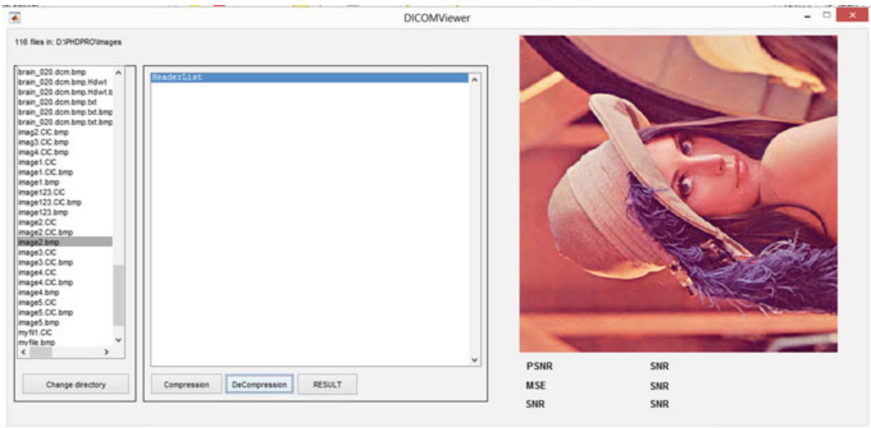


Fig. 10 Graphical user interface of image2.BMP for decompression

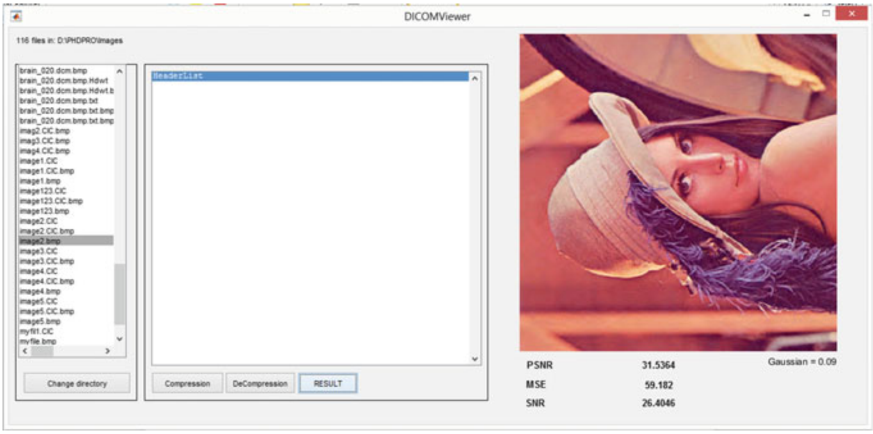


Fig. 11 Graphical user interface of image2.BMP for PSNR, MSE, SNR

of zeros in the transformation matrix. The EZSD algorithm used to remove zeros and at the same time it converts a high-frequency sub-bands into an array. Last, the arithmetic coding mapped entire sequence of symbols into a single codeword. It gives high encoding capacity.

Acknowledgements Authors thank Dr. S. V. Dudal, HOD, Department of Applied Electronics, SGBA University, Amravati and Maharashtra, India for providing all kind of facilities and support.

References

1. Malacara D (2011) Color vision and colorimetric: theory and applications. Press monograph, SPIE. http://books.google.com.br/books?id=xDU4YgEACAAJ
2. Frery AC, Perciano T (2013) Introduction to image processing using R. 21 Springer briefs in computer science. https://doi.org/10.1007/978-1-4471-4950-7-2
3. Gupta P, Bansal V (2015) The run length encoding for RGB images. Int J Converg Technol Manage 1(1). ISSN: 2455-7528
4. Xijun Y, Lei J (2015) A method of lossy compression for RGB565 format true color image. Int J Signal Process Image Process Pattern Recognit 8(5):279–288. http://dx.doi.org/10.14257/ijsip.2015.8.5.29. ISSN: 2005-4254
5. Boucetta A, Melkemi KE (2012) DWT based-approach for color image compression using genetic algorithm. In: International conference on image and signal processing, ICISP 2012: image and signal processing. LNCS 7340, pp 476–484
6. Sangwine SJ, Horne RE (1998) The colour image processing handbook, 1st edn. Chapman & Hall
7. Gonzalez RC, Woods RE (2001) Digital image processing. Addison Wesley Publishing Company, Reading
8. Sayood K (2000) Introduction to data compression, 2nd edn. Academic, Morgan Kaufman Publishers
9. Siddeq MM (2012) Using two level DWT with limited sequential search algorithm for image compression. J Signal Inf Process 3:51–62. https://doi.org/10.4236/jsip.2012

10. Naveen Kumar R, Jagadale BN, Sandeepa KS (2016) Use of optimal threshold and T-Matrix coding in discrete wavelet transform for Image compression. In: 3rd international conference on electronics and communication systems (ICECS 2016), IEEE Explore. 978-1-4673-7832-1/16
11. Buela Divya G, Krupa Swaroopa Rani K. Implementation of image compression using hybrid DWT–DCT algorithms, Int J Modern Trend Eng Res. ISSN: 2393-8161
12. Sangwine SJ, Horne RE (1998) The colour image processing handbook, 1st edn. Chapman & Hall
13. Sonka M, Halva V, Boyle T (1999) Image processing analysis and machine vision, 2nd edn. Brooks/Cole Publishing Company
14. Frery AC, Perciano T (2013) Color representation, introduction to image processing using R. 21 Springer briefs in computer science. https://doi.org/10.1007/978-1-4471-4950-7
15. Gersho A, Gray RM (1992) Vector quantization and signal compression, Boston. Kluwer Academic Publishers, MA
16. Suma S, Sridhar V (2014) A review of effective techniques of compression in medical image processing. Int J Comput Appl (0975–8887) 97(6)
17. Siddeq MM (2012) Using two level DWT with limited sequential search algorithm for image compression. J Signal Inf Process 3:51–62

Application of Neural Networks in Image Processing

Ishan Raina, Chirag Chandnani and Mani Roja Edinburgh

Abstract In today's world, the need for computers and their logics is critical for the development of any system. This paper is for those people who have little or no knowledge about Artificial Neural Networks (ANNs). Various advances have been made in creating intelligent systems, some enlivened by biological neural networks. Analysts from numerous logical orders are outlining artificial neural networks (ANNs) to take care of a variety of problems in design recognition, prediction, optimization, associative memory, and control. And will be discussing some of the algorithms of ANNs such as image segmentation using edge detection, image enhancement using multiscale retinex and technique combining sharpening and noise reduction and some practical applications such as diagnosing liver disease and face detection systems.

1 Introduction

ANNs are processing devices (paradigm or real equipment) that are loosely modeled after the neuronal structure of the mammalian cerebral cortex yet on considerably littler scales. A huge ANN may have hundreds or thousands of processor units, though a mammalian mind has billions of neurons with a relating increment in extent of their general connection and emanant conduct. Despite the fact that ANN analysts are for the most part not worried about whether their systems precisely take after organic frameworks, some have. For instance, specialists have precisely mimicked the capacity of the retina and displayed the eye rather well.

Neural systems are regularly composed in layers. Layers are comprised of various interconnected "hubs" which contain an "initiation work". Examples are introduced to the system by means of the "information layer", which conveys to at least one "hidden layers" the place the real preparing is done through an arrangement of weighted "associations". The shrouded layers at that point connect to a "yield layer" where the appropriate response is yield as demonstrated above. Modern computerized

I. Raina · C. Chandnani (✉) · M. R. Edinburgh
Thadomal Shahani Engineering College, Mumbai, India
e-mail: chandnanichirag512@gmail.com

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_48

Table 1 Comparison between VN and biological computer [1]

	Von Neumann computer	Biological computer
Processor	Complex High speed One or a few	Simple Low speed A large number
Memory	Separate from a processor Localized Noncontent and addressable	Integrated into processor Distributed Content addressable
Computing	Centralized Sequential Stored programs	Distributed Parallel Self-learning
Expertise	Numerical and symbolic manipulations	Perceptual problems
Reliability	Very vulnerable	Robust
Operating environment	Well-defined Well-constrained	Poorly-defined Unconstrained

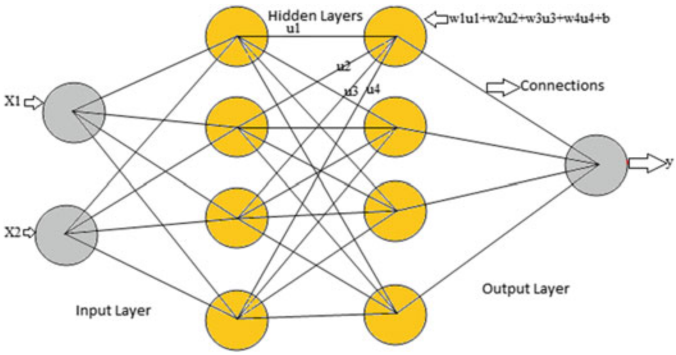


Fig. 1 The structure of a basic neural network

PCs beat people in the area of numeric calculation and related image control. In any case, people can easily take care of complex perceptual issues (like perceiving a man in a group from a minor look at his face) at such a rapid and degree as to predominate the world's fastest PC. Why is there such a surprising contrast in their execution? The organic neural framework design is totally not quite the same as the von Neumann engineering (see Table 1). This distinction essentially influences the sort of capacities each computational model can best perform (Fig. 1).

Various endeavors to create "savvy" programs in light of von Neumann's brought together engineering have not brought about broadly useful canny projects. Propelled by natural neural systems, ANNs are greatly parallel figuring frameworks comprising of an amazingly extensive number of straightforward processors with numerous interconnections. ANN models endeavor to utilize some "organizational" standards accepted to be utilized as a part of the human.

Moeled loosely on the human cerebrum, a neural net comprises of thousands or even a large number of basic preparing hubs that are thickly interconnected. The vast majority of the present neural nets are sorted out into layers of hubs, and they are "sustain forward," implying that information travels through them in just a single heading. An individual hub may be associated with a few hubs in the layer underneath it, from which it gets information, and a few hubs in the layer above it, to which it sends information.

To every one of its approaching associations, a hub will appoint a number known as a "weight." When the system is dynamic, the hub gets an alternate information thing—an alternate number—over every one of its associations and increases it by the related weight. It at that point includes the subsequent items together, yielding a solitary number. On the off chance that number is beneath a limit esteem, the hub passes no information to the following layer. On the off chance that the number surpasses the limit esteem, the hub "fires," which in the present neural nets by and large means sending the number—the entirety of the weighted information sources—along all its active associations.

At the point when a neural net is being prepared, the greater part of its weights and limits are at first set to arbitrary esteems. Preparing information is bolstered to the base layer—the information layer—and it goes through the succeeding layers, getting duplicated and included in complex courses, until the point that it at last arrives, fundamentally changed, at the yield layer. Amid preparing, the weights and limits are ceaselessly balanced until the point when preparing information with similar names reliably yield comparable yields.

The neural framework learns by altering its weights and slant (edge) iteratively to yield needed yield. These are moreover called free parameters. For making sense of how to happen, the neural framework is arranged first. The readiness is performed using described arrangement of standards generally called the training algorithm.

Popular Learning Algorithms used in Neural Network are as follows [1]:

Gradient Descent: This is the simplest training algorithm used in case of supervised training model. In case, the actual output is different from target output, the difference or error is find out. The gradient descent algorithm changes the weights of the network in such a manner to minimize this error.

Back Propagation: It is an extension of gradient-based delta learning rule. Here, after finding an error (the difference between desired and target), the error is propagated backward from output layer to the input layer via hidden layer. It is used in case of multilayer neural network.

1.1 Types of Learning Neural Network

Supervised Learning: In Supervised Learning, the training data as the initial input to the input layers and the required output is known to us. Weights are balanced until the point when output gives the required value.

Unsupervised Learning: The input information is utilized to prepare the system whose yield is known. The system characterizes the information and modifies the weight by feature extraction in input information.
Reinforcement Learning: Here the estimation of the yield is obscure, yet the system gives the feedback whether the yield is correct or off-base.
Offline Learning: The alteration of the weight vector and threshold is done simply after all the preparation set is displayed to the system.
Online Learning: The change of the weight and threshold is done. Subsequently, the every preparation sample is shown to the system.

1.2 Types of Data Sets Used in ANNs

Training Set: An arrangement of cases utilized for discovering that is to fit the parameters (i.e., weights) of the system. One Epoch contains one full preparing cycle on the training set.
Validation Set: An arrangement of cases used to tune the parameters (i.e., architecture) of the system. For instance to pick the quantity of hidden units in a neural network.
Test Set: An arrangement of cases utilized just to evaluate the performance of a completely determined system or to apply effectively in anticipating yield whose information is known.

*1.3 Different Uses of ANNs**

Classification: A neural system can be prepared to characterize given pattern or informational index into predefined class. It utilizes feed-forward systems.
Prediction: A neural system can be prepared to create yields that are expected from given info.
Clustering: The Neural system can be utilized to recognize a distinct component of the information and classify them into various classes with no earlier learning of the information.

2 Applications

2.1 Image Enhancement Using Multiscale Retinex

The retinex is a human recognition based image processing algorithm which gives shading steadiness and dynamic range compression. There are two kinds of retinex algorithms to be specific single scale retinex (SSR) and multiscale retinex (MSR). The MSR can be minimalistically composed as

$$F_i(x, y) = \sum_{n=1}^{N} W_n * \{\log[S_i(x, y)] - \log[S_i(x, y) * M_n(x, y)]\}, \tag{1}$$

where [2] where $F_i(x, y)$ is MSR and $S_i(x, y)$ is SSR the subscripts: $i\ E\ R, G, B$ speak to the three color bands, N is the quantity of scales being utilized, and W_n are the weighting factors for the scales. The $M_n(x, y)$ are the encompass function given by

$$M_n(x, y) = K_n \mathrm{e}^{[-(x^2+y^2)/\sigma_n^2]} \tag{2}$$

Rahman et al. [2] where σ_n the will be the standard deviations of the Gaussian distribution that decide the scale. The greatness of the scale decides the sort of data that the retinex gives: littler scales giving more unique range compression, and bigger scales giving more shading steadiness. The I (are selected so that [2]).

$$\iint F(x, y)\mathrm{d}x\mathrm{d}y = 1 \tag{3}$$

As expressed over, the MSR still experiences turning gray out of uniform zones much as the SSR did. The favorable position that the MSR has over the SSR is in the mix of scales which give both dynamic range compression and tonal version in the meantime. The general consequence of the use of the MSR is still more saturated than human perception, giving the last picture a washed-out appearance, however, it saves the vast majority of the detail in the scene. This turning gray of territories of consistent intensity happens in light of the fact that the retinex handling enhances each color band as a component of its encompass.

2.2 Image Enhancement Technique Combining Sharpening and Noise Reduction [3]

Another way to deal with contrast enhancement of picture information is displayed. The proposed strategy depends on a numerous yield framework that adopts fuzzy models with a specific end goal to keep the noise increase during the sharpening of the picture as points of interest. Key highlights of the proposed strategy are preferable execution over accessible techniques in the upgrade of pictures adulterated by Gaussian noise and no confounded tuning of fuzzy set parameters.

Enhancing the nature of sensor information is a key issue in picture based instrumentation. To be sure, preprocessing strategies can play an extremely important part in expanding the precision of consequent errands, for example, parameter estimation and object recognition. In this regard, contrast enhancement is frequently fundamental keeping in mind the end goal to highlight essential features embedded within the picture information. The improvement of noisy information, in any case, is an extremely basic process in light of the fact that the honing activity can altogether increase the noise. Distinctive systems have been proposed in the writing. They generally use the unsharp masking approach.

In the outstanding straight unsharp masking technique, a small amount of the high-pass filtered version of the input picture is added to the first information keeping in mind the end goal to get an enhanced picture. The reception of the high-pass channel, be that as it may, makes this basic strategy exceptionally sensitive to noise and frequently unacceptable for genuine applications. More powerful strategies utilize a nonlinear channel rather than the high-pass straight administrator with a specific end goal to accomplish a tradeoff between noise attenuation and edge enhancement. In this regard, an exceptionally intriguing class of nonlinear strategies is spoken to by polynomial unsharp masking methods, such as the Teager-based operator and the effective cubic unsharp masking strategy. In an alternate class of nonlinear methodologies, a weighted middle replaces the high-pass linear filter.

2.3 Image Segmentation Using Edge Detection

Image Segmentation is the way toward partitioning a computerized picture into numerous locales or sets of pixels. As a matter of fact, partitions are diverse objects in picture which have a similar texture or color. The outcome of image segmentation is an arrangement of areas that all in all cover the whole picture or an arrangement of forms extricated from the picture [4]. The majority of the pixels in a region are comparable with regard to some characteristic or computed property, such as color, intensity, or texture. Nearby areas are fundamentally different regarding the same attributes. Edge detection is a standout amongst the most as often as possible utilized procedures in computerized picture processing. The limits of object surfaces in a scene regularly prompt arranged confined changes in intensities of a picture, called edges. This perception joined with an accepted way of thinking that edge recognition is the initial phase in image segmentation and has powered a long inquiry for a decent edge detection paradigm to use in image processing. This search has constituted a foremost region of research in low-level vision and has prompted a constant flow of edge detection paradigm distributed in the image processing journals in the course of the most recent two decades. Indeed, even as of late, new edge detection paradigms are published every year.

Artificial neural systems (ANN) are generally connected for pattern recognition. Their handling potential and nonlinear attributes are utilized for clustering. Self-organization of Kohonen Feature Map (SOFM) network is an intense apparatus for grouping. Ji and Park proposed a paradigm for watershed segmentation in light of SOM [4]. This strategy finds the watershed segmentation of luminance part of color picture. The technique can be clarified as takes after. It comprises of two free neural systems one each for saturation and intensity planes. In edge location

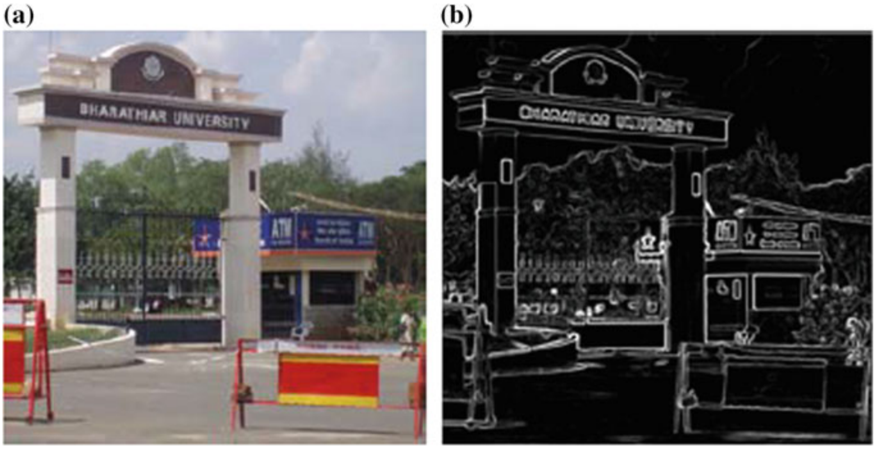


Fig. 2 **a** Original image [4]; **b** segmented image [4]

process, Initialize the synaptic weights of the system, to little, unique, arbitrary numbers at iteration, draw an example output from the information set. Locate the best coordinating (winning) neuron at this iteration utilizing the minimum distance Euclidean criterion. Continue this procedure until the synaptic weights__ achieve their Steady state values. This results in Image Segmentation (Fig. 2).

2.4 Diagnose of Disease Using Ultrasound Liver Images

Ultrasound imaging methodology is very well known and most generally utilized methodology for envisioning and concentrate the liver for any disease conditions without making any agony or distress the patient. Ultrasound liver imaging is generally utilized because of its noninvasive nature and ease when contrasted with other imaging modalities. The diagnosis of different illnesses is performed based on different picture features, for example, the echogenicity, legion shape, and echotexture. Liver imaging is a standout amongst other strategies of early detection of liver infections and early recognition is imperative since it spares patients from advance afflictions, for example, enlarged stomach filled with ascites fluid, bleeding varices, and encephalopathy or sometimes jaundice [5]. Liver sickness conditions, for example, liver disease conditions such as fatty liver, cirrhosis, and hepatomegaly are known for delivering unmistakable reverberate designs during US imaging, however, the pictures are likewise known to be visually challenging for interpreting them because of their imaging artifacts and speckle noise. Because of it, the sonographers need to depend upon additional pathological tests.

In this method image acquired are first processed using preprocessing techniques like cropping, edge detection then image processing is done in which feature extrac-

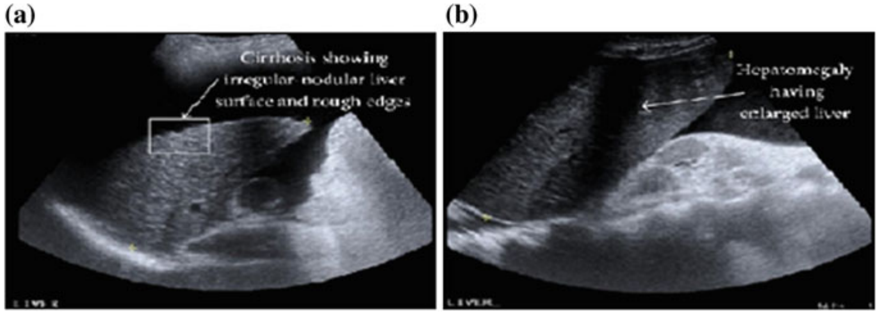


Fig. 3 **a** Cirrhosis liver [5]; **b** hepatomegaly liver [5]

tion is done after this final step is to select the feature which distinguishes the livers from each other. This is implemented using a ANN as explained below.

The backpropagation algorithm was picked because of the system's capacity to learn and store tremendous measures of mapping relations of information yield demonstrate without the requirement for earlier disclosure of mathematical conditions relating to these mapping relations. The algorithm likewise manages the system's weight and edge esteems so as to acquire minimum error sum of square. The outlined neural system classifier utilizes a two-layer feed-forward backpropagation network. Two-layer feed-forward network can be best characterized as a system with sigmoid hidden up and yield neurons.

To train the system, the input information and target information should be fed into the system. The system at that point partitions the input test information into three unique examples, which are training, validation, and testing samples. The training samples are utilized to prepare the system, and the system is balanced by its error. The validation samples are utilized to quantify organize generalization and to halt training when generalization stops improving. Testing samples are then used to give an autonomous measure of the system execution during and after training. In the event that the error of the system is still extensive, the system can be retrained back as to get more precise and proficient outcome. This is the way it is used to diagnose liver diseases using ANNs (Fig. 3).

2.5 Neural Network-Based Face Detection System

A neural network-based algorithm to identify upright, frontal perspectives of faces in grayscale images. The algorithm works by applying at least one neural networks straightforwardly to parts of the information picture and arbitrating their outcomes. Each system is prepared to yield the presence or absence of a face. The algorithms and training techniques are intended to be general, with little customization for faces.

This framework works in two phases.

- **Stage One: A Neural Network-Based Filter**

To detect faces anyplace in the input, the filter is connected at each area in the picture. To detect faces bigger than the window estimate, the information picture is over and over lessened in measure (by subsampling), and the filter is connected at each size. This filter must have some invariance to position and scale. The measure of invariance decides the quantity of scales and positions at which it must be connected.

- **Stage Two: Merging Overlapping Detections and Arbitration**

Two methodologies to enhance the unwavering quality of the detector: merging overlapping detections from a single network and arbitrating among multiple networks.

Most faces are detected at numerous adjacent positions or scales, while false discoveries frequently happen with less consistency. This perception prompts a heuristic which can dispense with numerous false detections. For every area and scale, the quantity of detections inside a predetermined neighborhood of that area can be tallied. In the event that the number is over a limit, at that point that area is delegated a face. The centroid of the close-by detections characterizes the area of the detections result, accordingly falling numerous detections. This is alluded to as "Thresholding" [6].

In the event that a specific area is effectively detected as a face, at that point all other detection areas which cover it are probably going to be error and can in this manner be eliminated. With the higher number of detections inside a small neighborhood and eliminate areas with less detections. This heuristic is called "overlap elimination" [6].

Arbitration Among Multiple Networks.

To additionally lessen the quantity of false positives, we can apply numerous systems and arbitrate between their yields to deliver an final decision. Each system is prepared in a comparable way, however with arbitrary beginning weights, irregular introductory nonface pictures, and permutations of the order of presentation of the scenery images. As will be found in the following segment, the detection and false-positive rates of the individual systems will be very close. Be that as it may, as a result of various preparing conditions and due to self-choice of negative preparing illustrations, the systems will have diverse predispositions and will make distinctive errors. One approach to consolidate two such pyramids is by ANDing them. This technique flags a detection just if the two systems identify a face at decisively a similar scale and position. Because of the diverse predispositions of the individual systems, they will once in a while concur on a error in detection of a face. This enables ANDing to take out most false detections from the given image in the input data.

3 Conclusion

In this paper, a brief information is mentioned on ANNs, types of ANN's that are currently being used in the world, different types of learning algorithm. The paper discusses mainly on different algorithms that are currently being used for various application of image processing such as image enhancement and Image Segmentation. Also, some real-life applications of these image processing techniques are mentioned in the article such as image processing for identification of a disease in the medical field and face detection algorithm.

References

1. Jain AK, Mao J, Mohiuddin KM. Artificial neural networks: a tutorial
2. Rahman Z, Jobson DJ, Woodell GA. Multi-scale retinex for color image enhancement
3. Russo F. An image enhancement technique combining sharpening and noise reduction
4. Senthilkumaran N, Rajesh R. Edge detection techniques for image segmentation
5. Systems Biomedicine Division, India Research Advisory Council, Haffkine Institute for Training Research and Testing, Parel, Mumbai, Maharashtra
6. Rowley HA, Baluja S, Kanade T. Neural network-based face detection

Region of Interest (ROI) Based Image Encryption with Sine Map and Lorenz System

Veeramalai Sankaradass, P. Murali and M. Tholkapiyan

Abstract In this research work, ROI-based grayscale image encryption with chaos is proposed. First, ROI areas are identified using Sobel edge detection operator and categorized into important and unimportant regions based on number of edges present in the particular block. Next, the important regions are encrypted with Lorenz system (both confusion and diffusion process) and unimportant regions are encrypted using Sine map. Finally, the entire image is shuffled by Lorenz system with new initial conditions to get final encrypted image. The significant advantage of this research work is that important and unimportant regions are encrypted separately with different chaos equation and system which increases the security of the image and also the end user can vary the important regions depends upon the requirements. The experimental results show that the proposed encryption approach provides better results for different cryptographic attacks.

1 Introduction

Security of data including image, video, text, audio and multimedia is gaining much attention due to vast development of communication technologies. During transmission over public networks as well as private networks, hackers and attackers may steal that data for various purposes such as modification, fabrication, etc. Therefore, data security has become an important issue during transmission as well as

V. Sankaradass · P. Murali (✉)
Department of Computer Science and Engineering, Vel Tech High Tech Dr. Rangarajan Dr. Sakunthala Engineering College, Avadi, Chennai, India
e-mail: pmuraliphd@gmail.com

V. Sankaradass
e-mail: veera2000uk@gmail.com

M. Tholkapiyan
Department of Civil Engineering, Vel Tech High Tech Dr. Rangarajan Dr. Sakunthala Engineering College, Avadi, Chennai, India
e-mail: m.tholkapiyan@gmail.com

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_49

storage. Recently, lots of images are transmitted and uploaded to the Internet and image encryption is serious issue. Traditional encryption algorithms such as Data Encryption Standard (DES), Advanced Encryption Standard (AES), etc., are mainly developed for text data and not suitable for images due to bulk volume of data, high correlation among pixels and intrinsic features of image [1–5]. By contrast, chaos-based image encryption algorithms have attracted many researchers due to high sensitivity to initial conditions, ergodicity, pseudorandomness and periodicity [6–9].

2 Literature Survey

In this section, related works are presented. Nowadays, lots of image encryption schemes based on chaos are presented. Ye et al. [10] presented image encryption algorithm based on generalized Arnold's map. In this work, first, total circular function is used in the permutation stage and in the diffusion stage, double diffusion process is applied with keystream. Image encryption with combined confusion and diffusion is designed in [1]. The spatiotemporal chaos is utilized to shuffle the pixels as well as diffusion and also used for generating random numbers.

In [11], image encryption scheme based on logistic map and DNA is developed. First, 2-D logistic map is employed to shuffle the pixels in the image then, DNA encoding is applied to encode the shuffled pixels. Finally, DNA addition and subtraction are performed to obtain the ciphered image. Region of Interest (ROI) based image encryption is gaining popularity in recent days. In [12], selective image encryption based on orthogonal polynomials transformation (OPT) and chaos for low power devices are proposed. In this work, ROIs are identified with orthogonal polynomials and important regions are encrypted in bit level with chaos. Next, unimportant regions are encrypted in the OPT domain with chaos. Finally, square-wave SFC is applied to obtain the final encrypted image. Tanja et al. [13] presented combinational domain image encryption. First, they identified significant and insignificant areas in the image using Prewitt edge detection. Next, the identified significant areas are encrypted in the spatial domain with chaos and insignificant areas are encrypted in the wavelet domain. Finally, they applied reverse diffusion process to get the final encrypted image.

Motivated by existing image encryption methods, in this research work, we propose novel image encryption scheme based on ROI and chaos. First, the Sobel edge detection operator is applied to plain image and categorized into important and unimportant regions based on number of edges. Then, important region blocks are encrypted with Lorenz system and unimportant region blocks are encrypted with Sine map. Finally, overall shuffling is performed using random sequences from Lorenz system. The main feature of the proposed work is separate encryption of important and unimportant regions which increases the security of the image.

The remainder of this paper is organized as follows: In Sect. 3, details of Lorenz system and sine map are presented. Section 4 describes the proposed cryptosystem. The experiments, results and discussions carried out to show the effectiveness of the proposed system are presented in Sect. 5 and finally conclusions are drawn in Sect. 6.

3 Chaotic System and Map

In the proposed research work, Lorenz system and sine map are used. The Lorenz system is used to encrypt the important regions in the image as well as shuffling of entire image. The sine map is used to encrypt unimportant regions.

3.1 Lorenz System

The Lorenz system with x, y and z as state variables is mathematically defined as

$$\begin{aligned} \dot{x} &= -a(x - y) \\ \dot{y} &= -xz + bx - y \\ \dot{z} &= xy - cz \end{aligned} \tag{1}$$

where a, b and c are control parameters and when $a = 10$, $b = 28$ and $c = 8/3$, the system is in chaotic state.

3.2 Sine Map

In this proposed work, sine map is used to generate random values and hence to construct random window in the random artificial image. The 1D sine map, $x(i)$ is mathematically defined as

$$x(i) = \lambda * \text{sine}(\pi * (i - 1)) \tag{2}$$

where $\lambda = 0.99$ and $\pi = 3.14$.

Having described the mathematical preliminaries, the proposed cryptosystem is presented in the next section.

4 Proposed Cryptosystem

4.1 Proposed Encryption Process

The proposed encryption method starts with identification of ROI areas in the image. Generally, image contains various areas like edges, smooth regions, etc. The proposed research work applies Sobel edge detection operator on the input image of size $(M \times N)$ and obtains the edge detected binary image. Next, the binary image is divided into $(k \times k)$ non-overlapping blocks where $k < M, N$ and number of edges present in the each block is counted. Then, based on threshold value, T from the user, blocks with edges are separated as important and unimportant regions where number of edges are greater than in important regions and less than in unimportant regions. Also, the identified important and unimportant regions are termed as 1 and 0 that is act as secret key in the decryption side. After that, important blocks are encrypted with Lorenz system. Before encryption, the sine map is iterated and their values are digitized to 1, 2 and 3. Then, Lorenz system is iterated $(k \times k)$ times and generate the sequences. Based on sine map values, one of the sequence is selected from Lorenz system and sorted. Finally, index values are obtained from sorted sequence to shuffle the block. Next, exclusive-OR (XOR) operation is performed between generated sequence (digitized values) from Lorenz systems and shuffled values. The above process is repeated for all important blocks with new initial conditions.

Next, Similar to important blocks encryption, unimportant blocks are encrypted with sine map. Finally, Lorenz system is iterated $(M \times N)$ times and the index values are used to shuffle the entire image to produce the final encrypted image. The steps involved in the proposed encryption method is presented as an algorithm hereunder:

Algorithm

Input: Grayscale image of size $(M \times N)$
Output: Encrypted image of size $(M \times N)$
Begin
Step 1: Read plain image of size $(M \times N)$
Step 2: Apply Sobel edge detection operator on the plain image
Step 3: Divide the binary image obtained from Step 2 into $(k \times k)$ non-overlapping blocks where $k < M, N$
Step 4: Count the edge points cnt, in each block of the binary edge map
Step 5: Accept threshold T from user
Step 6: If the $cnt \geq T$, then the corresponding block is considered as important and termed as ROI; Else, it is considered as unimportant, non-ROI
Step 7: Identified blocks (important and unimportant) are numbered from left to right, 1 and 0
Step 8: Repeat the above steps for all blocks in the image
Step 9: Iterate the Sine map up to total number of important blocks and digitize the values to 1, 2 and 3

Step 9: Important Blocks Encryption
Step 9.1: Iterate Lorenz system ($k \times k$) times with new initial conditions
Step 9.2: Based on digitized sequence, select one of the sequence from Lorenz system and sort the sequence into S_x and obtain the respective index I_x. It can be represented as

$$I_x[i] \leftarrow \left[S_x, X^l\right]$$

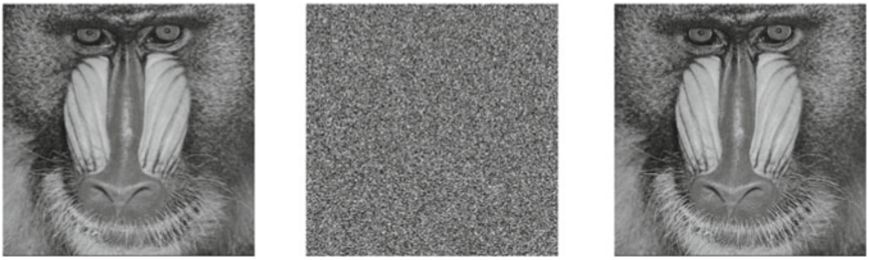
Step 9.3: Shuffle the block using I_x and obtain the shuffled blocks
Step 9.4: Consider the same sequence and digitize the values between 0 and 255
Step 9.5: Diffuse the shuffled block by performing Exclusive-OR (XOR) between values from Step 9.4 and shuffled block
Step 9.6: Repeat Step 9.1 to 9.5 for all important blocks
Step 10: Unimportant blocks encryption using Sine map that is similar to important blocks encryption
Step 11: Iterate Lorenz system ($M \times N$) times with new initial conditions and select one of the sequence and obtain index values from sorted sequence
Step 12: Perform shuffling operation based on index values for the entire image and obtain the final encrypted image
End

4.2 Proposed Decryption Process

The main aim of the decryption process is to retrieve the original image from encrypted image. The proposed decryption process is the exact reverse of encryption process with same secret keys.

5 Experiment and Security Analysis

The proposed encryption scheme is experimented with more than 100 benchmark images of size (512×512). First, the edge detected image is divided into (32×32) non-overlapping blocks. Then, Sine map is iterated to generate values 1, 2 and 3 and the corresponding keys are $x = 0.1346$, $\lambda = 0.99$ and $\pi = 3.14$. For important blocks encryption, the lorenz system values are $a = 10, b = 28, c = \frac{8}{4}, x_1 = 10.0, y_1 = 15.0$ and $z_1 = 45.0$. Similarly, for unimportant blocks encryption, the sine map values are $x = 0.16746$, $\lambda = 0.99$ and $\pi = 3.14$. Finally, overall shuffling is performed with Lorenz system and the corresponding values are $a = 10, b = 28, c = \frac{8}{4}, x_1 = 10.0, y_1 = 17.0$ and $z_1 = 45.0$. The proposed encryption method encrypts the original mandrill image with above keys. Figure 1 shows the encryption and decryption results of Mandrill image.



(a) Original Mandrill Image (b) Encrypted Mandrill Image (c) Decrypted Mandrill Image

Fig. 1 Encryption and decryption results of proposed technique

The proposed encryption technique is analyzed with various performance measures and the corresponding results are presented in the following subsections.

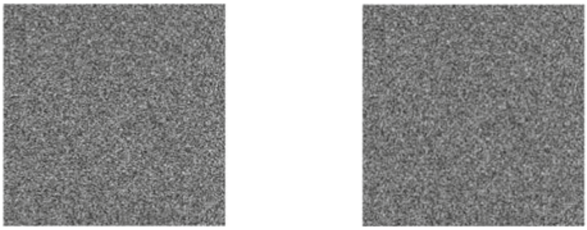
5.1 Key Sensitivity Analysis

The encrypted image is sensitive with secret keys values. To test the key sensitivity of the proposed cryptosystem, first, the original grey image is encrypted with above said key 1 ($x = 0.1346, \lambda = 0.99, \pi = 3.14, a = 10, b = 28, c = \frac{8}{4}, x_1 = 10.0, y_1 = 15.0, z_1 = 45.0, x = 0.16746, \lambda = 0.99, \pi = 3.14, a = 10, b = 28, c = \frac{8}{4}, x_1 = 10.0, y_1 = 17.0$ and $z_1 = 45.0$). Then, the original key 1 is modified as key 2. The modified key values are ($x = 0.1347, \lambda = 0.99, \pi = 3.14, a = 10, b = 28, c = \frac{8}{4}$, $x_1 = 10.0, y_1 = 15.0, z_1 = 45.0, x = 0.16746, \lambda = 0.99, \pi = 3.14, a = 10, b = 28, c = \frac{8}{4}, x_1 = 10.0, y_1 = 17.0$ and $z_1 = 45.0$). Again, the key 1 is slightly modified as key 3 and the corresponding values are ($x = 0.1346, \lambda = 0.99, \pi = 3.14, a = 10, b = 28, c = \frac{8}{4}, x_1 = 10.0, y_1 = 15.0, z_1 = 45.0, x = 0.16746, \lambda = 0.99$, $\pi = 3.14, a = 10, b = 28, c = \frac{8}{4}, x_1 = 10.0, y_1 = 17.000001$ and $z_1 = 45.0$). Now the encrypted image is decrypted with key 2 and key 3 and the corresponding results are shown in Fig. 2a, b.

From Fig. 2, the proposed encryption technique is highly sensitive to secret keys. So, small change in the key structure will not produce original image in the proposed method.

5.2 Histogram Analysis

Figure 3a, c shows the histogram of original Lena and Mandrill image. Similarly, Fig. 3b, d shows the corresponding encrypted Lena and Mandrill image, respectively.



(a) Decrypted Mandrill image with Key2 (b) Decrypted Mandrill image with Key3

Fig. 2 Decrypted images with wrong keys

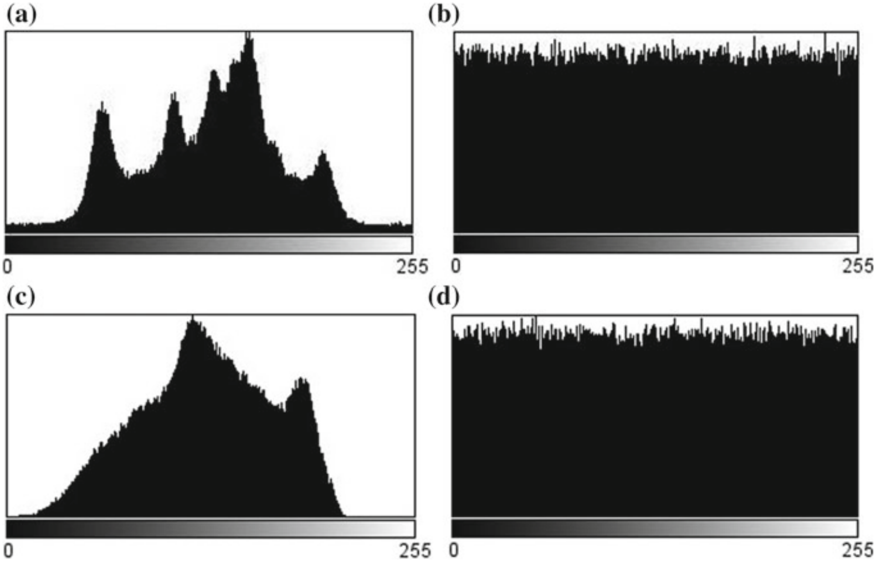


Fig. 3 Histogram analysis results **a** original Lena image, **b** encrypted Lena Image, **c** original Mandrill image, **d** encrypted Mandrill image

From the figures, histogram of encrypted images is flat and one cannot identify the different images. Therefore, the proposed encryption method is successful against statistical attack.

5.3 Differential Attack

To find out the relationship between original image and encrypted image, the attacker encrypts the original image before and after changing and comparing those images.

It is called differential attack and it is measured by the number of pixel rate change (NPCR) and the unified average changing intensity (UACI).

5.3.1 Number of Pixel Change Rate (NPCR)

The NPCR is used to measure the number of pixels indifference of a component between two images.

$$\text{NPCR} = \frac{\sum_{i,j} D(i, j)}{W \times H} \times 100\% \tag{3}$$

$$\text{where } D(i, j) = \begin{cases} 1, & c_1(i, j) \neq c_2(i, j) \\ 0, & \text{otherwise} \end{cases}$$

Here W and H are the width and height of the image, c_1 and c_2 are two cipher images with slightly modified keys, $c_1(i, j)$ and $c_2(i, j)$ are pixel values of c_1 and c_2 at position (i, j).

5.3.2 Unified Average Change Intensity (UACI)

UACI is used to measure the average intensity difference in component and can be defined as

$$\text{UACI} = \frac{1}{N}\left[\sum_{i,j} \frac{\left|c(i, j) - c'(i, j)\right|}{2^L - 2}\right] \times 100\% \tag{5}$$

The proposed cryptosystem is analyzed with NPCR and UACI and the values are presented in Table 1.

From Table 1, it shows that there is a small change in the original images will result in a significant change in the encrypted images. The table also implies that the proposed encryption technique successfully resists the differential attack.

Table 1 NPCR and UACI values between original Lena image and encrypted Lena image

Image	NPCR%	UACI%
Lena	99.87	31.59
Mandrill	99.64	31.68
House	99.39	31.62
Lake	99.44	31.57
Fruits	99.71	31.78

Table 2 Results of PSNR values obtained from proposed encryption scheme

Image	PSNR in dB
Lena	10.67
Mandrill	10.91
House	9.78
Lake	9.29
Fruits	10.22

5.4 Peak Signal-to-Noise Ratio (PSNR)

The performance of the proposed cryptosystem is evaluated objectively via the PSNR between the original image and the encrypted image. The mathematical formula for computation of PSNR is

$$\mathrm{PSNR} = 10\log_{10}\left(R^2/\mathrm{MSE}\right) \tag{6}$$

where R is maximum gray level and MSE is the mean square error given by

$$\mathrm{MSE} = \left(\sum_{m,n}[I_1(i,j) - I_2(i,j)]^2/(m \times n)\right)$$

$I_1(i, j)$ and $I_2(i, j)$ represent the intensity of the original and encrypted image pixel positions at (i, j), and m and n are the height and width of the image. The PSNR results of the proposed encryption scheme are tabulated in Table 2.

From Table 2, low PSNR values are produced for all images that means difficulty in retrieving the original image.

6 Conclusion

This research work presents ROI-based image encryption with Lorenz system and Sine map. First, the ROIs are identified using Sobel edge detection operator and important and unimportant regions are encrypted with Lorenz system and Sine map, respectively. The main advantage of the proposed encryption technique is encryption of ROI areas separately with different chaos which maximizes the security of the encrypted images. Extensive security analyses have been carried out to demonstrate the efficiency of the proposed system, suitable for practical applications.

References

1. Wang Y, Wong K-W, Liao X, Chen G (2011) A new chaos-based fast image encryption algorithm. Appl Soft Comput 11(1):514–522
2. Abdullah AH, Enayatifar R, Lee M (2012) A hybrid genetic algorithm and chaotic function model for image encryption. AEU Int J Electron Commun 66(10):806–816
3. Kalpana J, Murali P (2015) An improved color image encryption based on multiple DNA sequence operations with DNA synthetic image and chaos. Opt Int J Light Electron Opt 126(24):5703–5709
4. Afarin R, Mozaffari S (2013) Image encryption using genetic algorithm. In: 2013 8th Iranian conference on machine vision and image processing (MVIP), pp 441–445
5. Ayoup AM, Hussein AH, Attia MAA (2015) Efficient selective image encryption. Multimed Tools Appl
6. Babaei M (2013) A novel text and image encryption method based on chaos theory and DNA computing. Nat Comput 12(1):101–107
7. Bhatnagar G, Jonathan Wu QM (2012) Selective image encryption based on pixels of interest and singular value decomposition. Digit Signal Process 22(4):648–663
8. Bigdeli N, Farid Y, Afshar K (2012) A novel image encryption/decryption scheme based on chaotic neural networks. Eng Appl Artif Intell 25(4):753–765
9. Chai X, Gan Z, Yuan K, Chen Y, Liu X (2017) A novel image encryption scheme based on DNA sequence operations and chaotic systems. Neural Comput Appl
10. Ye G, Wong K-W (2012) An efficient chaotic image encryption algorithm based on a generalized Arnold map. Nonlinear Dyn 69(4):2079–2087
11. Wang X-Y, Zhang Y-Q, Zhao Y-Y (2015) A novel image encryption scheme based on 2-D logistic map and DNA sequence operations. Nonlinear Dyn 82(3):1269–1280
12. Krishnamoorthi R, Murali P (2017) A selective image encryption based on square-wave shuffling with orthogonal polynomials transformation suitable for mobile devices. Multimed Tools Appl 76(1):1217–1246
13. Taneja N, Raman B, Gupta I (2012) Combinational domain encryption for still visual data. Multimed Tools Appl 59(3):775–793

Chaos-Based Color Image Encryption with DNA Operations

P. Murali, Veeramalai Sankaradass and M. Tholkapiyan

Abstract In this paper, an efficient yet simple color image encryption is proposed based on DNA sequence operation and chaotic maps for Android-based mobile devices. We designed our encryption algorithm based on such constraints. The main idea of the proposed method is different kind of DNA encoding for RGB channels and then Arnold's cat map shuffling is performed in RGB channel in the combined manner. Then DNA addition is performed to diffuse the values between the RGB channels itself. Finally, DNA decoding is performed to form the color encrypted image. The promising feature of the proposed encryption algorithm is simple operations which comfortably suites for low processing power computing devices like Android mobile phones. The analysis and experiment's results show the efficiency of the proposed encryption method.

1 Introduction

Nowadays, the communication between the people is drastically increased when compared to last decade. The development of many electronic devices like mobile phones, smartphones, etc., made communication much easier. So the people can use the Internet within mobile devices at any time and anywhere. However, the secrecy of data is a major issue. To protect data while transmission, encryption plays an important role. Among various encryption methods, image encryption is totally

P. Murali (✉) · V. Sankaradass
Department of Computer Science and Engineering, Vel Tech High Tech Dr. Rangarajan Dr. Sakunthala Engineering College, Avadi, Chennai, India
e-mail: pmuraliphd@gmail.com

V. Sankaradass
e-mail: veera2000uk@gmail.com

M. Tholkapiyan
Department of Civil Engineering, Vel Tech High Tech Dr. Rangarajan Dr. Sakunthala Engineering College, Avadi, Chennai, India
e-mail: m.tholkapiyan@gmail.com

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_50

different due to bulk capacity, redundancy, etc. The traditional encryption methods like DES, AES, etc., are unsuitable for multimedia encryption due to intrinsic nature of data [1–5]. So, the researchers developed various image encryption methods based on spatial domain and frequency domain. In these methods, the chaos-based image encryption is familiar due to high security, high sensitivity, large key space, etc. [3–5].

2 Literature Survey

In this section, related works are discussed. Recently, Kanso et al. has been proposed color image encryption algorithm, which uses 3D chaotic map [6]. In [7], chaotic neural networks based image encryption and decryption is introduced, which uses like Lorenz, Chua system, Lu system, Tent map, and Arnold's cat map. Liu et al. [8] presented bit level based color image encryption which Piecewise Linear Chaotic Map (PWLCM) and random numbers. The development of DNA computing, DNA cryptography is introduced and biological technology is used as a tool for implementing encryption, message hiding, etc. [9–12]. In [9, 10], initially, DNA encoding is applied to pixels, and diffusion is performed using DNA addition operation. Finally, DNA decoding is performed to construct color encrypted image. The main drawback is fixed DNA encoding and decoding. DNA subsequence operation is introduced in [11, 12].

The above-discussed DNA based image encryption algorithms have drawbacks of fixed DNA encoding and decoding rules which creates the loophole for attackers. In order to overcome the above shortcomings and motivated by DNA computing, we propose simple and efficient color image encryption scheme based on DNA sequence operations and chaotic maps in this paper. Also, we propose dynamic DNA encoding and decoding rules for RGB channels which increase the randomness and security of the encryption algorithm. The most important feature is diffusion operation, which is done between RGB channels. The remaining of this paper is organized as follows. Arnold's cat map and 1D Chebyshev map are introduced in Sect. 2. In Sect. 3, DNA sequences and DNA addition and subtraction are presented. In Sect. 4, the proposed encryption and decryption method is presented, and experiments and analysis are presented in Sect. 5. Finally, conclusion is drawn in Sect. 6.

3 Chaotic Maps

In our proposed algorithm, we use two kinds of chaotic maps: Arnold's cat map and 1D Chebyshev map. The cat map is used for shuffling purpose and Chebyshev map is used for generating the different kind of DNA encoding and decoding rules for RGB channels with different initial conditions.

3.1 Arnold's Cat Map

It is a two dimensional invertible map. It was discovered by Russian mathematician Vladimir. I. It is used to scramble the DNA encoded values in the proposed method. It is mathematically defined by

$$\begin{pmatrix} x_{n+1} \\ y_{n+1} \end{pmatrix} = \begin{pmatrix} 1 & p \\ q & pq+1 \end{pmatrix} \begin{pmatrix} x_n \\ y_n \end{pmatrix} \bmod 1 = A \begin{pmatrix} x_n \\ y_n \end{pmatrix} \bmod 1 \tag{1}$$

where $x \bmod 1$ refers the fractional part of x for any real number x, p, and q are cat map control parameters. The map is area-preserving since $\det|A| = 1$. An interesting property of Arnold's cat map is it will return into a state which is very close to the initial state after a few iterations. The cat map parameter p and q acts as secret key while decryption process.

3.2 1D Chebyshev Map

The k_{th} degree 1D polynomial can be mathematically defined as

$$T_k(x) = \cos(k \arccos x) \tag{2}$$

where $k = 1, 2, 3, \ldots, n$ and $x \in [-1, 1]$.

Here, 1D Chebyshev map is used to generate the encoding and decoding rule for RGB channels with different initial condition and the mathematical formula is defined as

$$\text{Rul}_i = f((\cos(k \arccos x) * 1000))\%8 \tag{3}$$

where $i = 1, 2, 3, \ldots, (M \times 4N)$ and $f()$ denotes the integer function.

4 Chaotic Maps

The DNA sequence consists of four nucleic bases A (adenine), C (cytosine), G (guanine), T (thymine), and can be represented ad 00, 01, 10 and 10. There are 24 encoding rules, but only eight rules are satisfying the Watson–Crick complement rule. The sequence is shown in Table 1. In the four bases, A and T are complementary. G and C are complementary.

The DNA sequence is used to encode the pixels. The pixel consists of 8 bits, and it is separated into four parts each consists of 2 bits, and one of the DNA rules is used to encode the separated four parts. For example, the pixel value is 189, and

Table 1 DNA sequence

	1	2	3	4	5	6	7	8
A	00	00	01	01	10	10	11	11
T	11	11	10	10	01	01	00	00
G	01	10	00	11	00	11	01	10
C	10	01	11	00	11	00	10	01

the binary value is 10111101. It is separated into four parts like 10, 11, 11 and 01 and DNA encoded with rule 5 and the DNA sequence becomes TGGA. To get the original binary sequence, rule 5 is applied if we use different DNA rule, we can get wrong DNA sequence.

4.1 DNA Addition and Subtraction Operation

DNA addition and subtraction operation is performed in binary level. The DNA addition and subtraction rule table are shown in Tables 2 and 3. In image encryption, these two operations are performed while diffusing the pixel's values. In Encryption, the two pixels are added (DNA addition) and DNA subtraction is performed during the decryption process to retrieve the original pixel value.

For example, two DNA sequences [AGCT] and [TAGG] and we applied DNA addition rule as shown in Table 2, and we get the DNA sequence [TGTA]. Similarly, we can obtain the original sequence [AGCT] by subtracting [TAGG] from [TGTA]. In our proposed encryption algorithm, we adopted DNA addition operation during the encryption process and DNA subtraction operation during the decryption process.

Table 2 DNA addition

+	A	G	C	T
A	A	G	C	T
G	G	C	T	A
C	C	T	A	G
T	T	A	G	C

Table 3 DNA subtraction

–	A	G	C	T
A	A	T	G	C
G	G	A	T	C
C	C	G	A	T
T	T	C	G	A

5 Proposed Encryption Technique

First, the original color image with size ($M \times N$) is separated into RGB channel and then converted into a bit format with size ($M \times 8N$). Next, we applied DNA encoding operation on RGB channel using Eq. 3 with different initial condition for every channel. So every pixel is differently encoded unlike DNA encoding. Then the entire encoded RGB channel is combined in vertical direction, and we applied Arnold's cat map to scramble the encoded values and it is separated into three equal parts with size ($M \times 4N$). DNA addition operation is performed between scrambled channels to diffuse the values furthermore. After that, DNA decoding operation is performed on the diffused channels with three different rules as described in Eq. 3, and every channel is converted to the pixel format. Finally, the entire three channels are combined to form the encrypted color image. These are presented as an algorithm in the following sections.

5.1 Proposed Encryption Algorithm

Input: Plain Image of size ($M \times N$)

Output: Encrypted Image of size ($M \times N$)

Begin

Step 1 Read an original color image of size ($M \times N$).

Step 2 Convert color image into R, G, and B channel with size ($M \times N$).

Step 3 Convert R, G and B channels into bit format with size ($M \times 8N$).

Step 4 Apply DNA encoding on RGB channel as described in Eq. 3. It is denoted as R_{en}, G_{en}, and B_{en} as and the size of every channel becomes ($M \times 4N$).

Step 5 Combine the encoded channels R_{en}, G_{en} and B_{en} in vertical direction with size ($M \times 3(4N)$). It is denoted as RGB_{com}.

Step 6 Initialize Arnold's cat map parameters p and q.

Step 7 Shuffle the RGB_{com_s} by Arnold's cat map with n times. It is denoted as RGB_{com_s}.

Step 8 Split the RGB_{com_s} into three separate RGB_{com} channels with size ($M \times 4N$). It is named as R_{en_s}, G_{en_s} and B_{en_s}.

Step 9 Perform the following DNA Addition operation on shuffled channels:
$R' = R_{en\_S} + B_{en\_s}$, $G' = G_{en\_S} + B_{en\_s}$, $B' = B_{en\_S} + R'$
where + denotes the DNA addition operation.

Step 10 Perform DNA decoding on channels R', G' and B' as described in Eq. 3. It is denoted as R_f, G_f and B_f with size ($M \times 8N$).

Step 11 Convert R_f, G_f and B_f into pixel format with size ($M \times N$). It is denoted as R_{pix}, G_{pix} and B_{pix}.

Step 12 Combine R_{pix}, G_{pix} and B_{pix} to get final encrypted color image.

End

5.2 Proposed Decryption Algorithm

The decryption process is just reverse of the encryption algorithm. In encryption, we performed DNA addition operation to diffuse the pixel values but in decryption, DNA subtraction operation is applied to get back the original values. Here, we present only DNA subtraction formula:

$$\mathrm{B_{en\_s}} = \mathrm{B}' - \mathrm{R}',\ \mathrm{G_{en\_s}} = \mathrm{G}' - \mathrm{B_{en\_s}},\ \mathrm{R_{en\_s}} = \mathrm{R}' - \mathrm{B_{en\_s}}$$

where denotes the DNA addition operation.

6 Experiments and Security Analysis

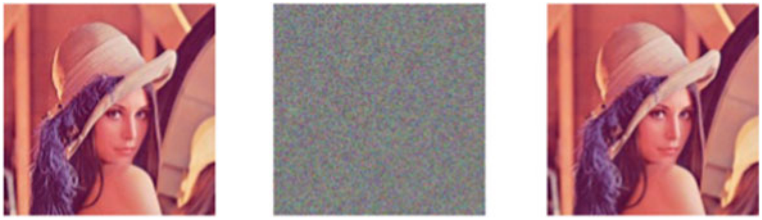
The proposed color image encryption algorithm is experimented with more than 100 color images which include standard image processing images with size of (512 × 512).

6.1 Encryption and Decryption

The proposed encryption algorithm uses the following keys to obtain the encrypted image. The key values are 0.000001, 2, 0.000011, 3, 0.000111, 4, 50, 49, 2, 0.000002, 5, 0.000022, 6, 0.000222, and 3. The values 0.000001, 2, 0.000011, 3, 0.000111, 4 in the key structure denotes the 1D Chebyshev map initial values and *k* values for RGB channel to generate DNA encoding rules and the values 50, 49, and 2 denotes Arnold's cat map parameter and number of iteration. Finally, the values 0.000002, 5, 0.000022, 6, 0.000222, and 3 denotes the 1D Chebyshev map initial values and k values for RGB channel to generate DNA decoding rules. For representation purpose, we show only Lena image with size of (512 × 512). Figure 1a shows the original color Lena image, and the encrypted color image is shown in Fig. 1b. The cipher image is decrypted successfully with abovementioned key, and the decrypted image is shown in Fig. 1c.

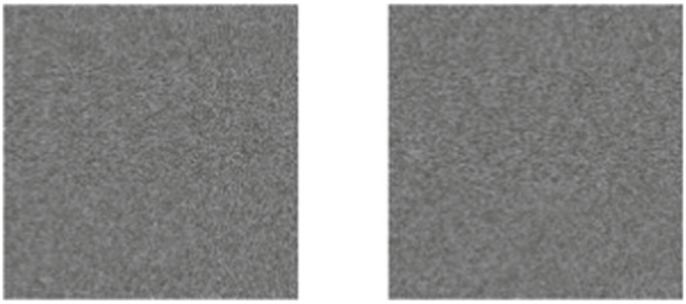
6.2 Key Sensitivity Analysis

For good cryptosystem, the encrypted image is sensitive with secret key values. To test the key sensitivity analysis, first we encrypt the original color image with key 1 (0.000001, 2, 0.000011, 3, 0.000111, 4, 50, 49, 2, 0.000002, 5, 0.000022, 6, 0.000222, and 3) and then we slightly modify the key 1, and it is denoted as key



(a) Original Lena image (b) Encrypted Lena image (c) Decrypted Lena image

Fig. 1 Encryption and decryption results of the proposed algorithm



(a) Decrypted Lena image with key2 (b) Decrypted Lena image with key3

Fig. 2 Decrypted color images with wrong keys

2. The modified key values are (0.000011, 2, 0.000011, 3, 0.000111, 4, 50, 49, 2, 0.000002, 5, 0.000022, 6, 0.000222, and 3). Again, the key1 is slightly modified as key 3 and the corresponding values are (0.000001, 2, 0.000011, 3, 0.000112, 4, 50, 49, 2, 0.000002, 5, 0.000022, 6, 0.000222, and 3). Now, the encrypted image is decrypted with key2 and key3, and the results are presented in Fig. 2.

It is clear from the results that the minor modification in the key structure will give a completely different output. So the proposed encryption algorithm is highly sensitive with the secret key.

6.3 Histogram Analysis

The histogram of the image shows the how pixels are distributed. The pixel distribution of the proposed encryption algorithm is tested in terms of histogram analysis. The results are presented in Fig. 3.

Figure 3a–c shows the pixel distribution of original Lena image R, G, and B channel and Fig. 3d–f shows the encrypted Lena's image R, G and B channel. It is

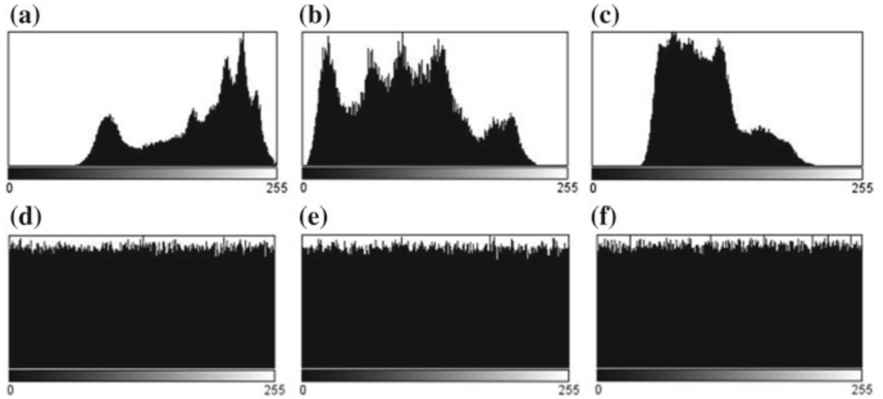


Fig. 3 Histogram analysis results, **a** original Lena image Red channel, **b** Green channel, **c** Blue channel, **d** encrypted Lena image Red channel, **e** Green channel, **f** Blue channel

evident that pixels in encrypted image are evenly arranged and does not give any information about the original image.

6.4 Differential Attack

The attacker usually slightly modifies the original image and then performs the encryption and compares the results after changing and before changing. It is called differential attack. It is measured by the number of pixel rate change (NPCR) and the unified average changing intensity (UACI).

6.4.1 Number of Pixel Change Rate (NPCR)

The NPCR is used to measure the number of pixels in difference of a component between two images.

$$\text{NPCR} = \frac{\sum_{i,j} D(i, j)}{W \times H} \times 100\% \tag{4}$$

where W and H is the width and height of the image and $D(i, j)$ is defined as

$$\text{where } D(i, j) = \begin{cases} 1, & c_1(i, j) \neq c_2(i, j) \\ 0, & \text{otherwise.} \end{cases} \tag{5}$$

Table 4 NPCR and UACI values between original Lena image and encrypted image

	NPCR (%)			UACI (%)		
Image	Red	Green	Blue	Red	Green	Blue
Lena	99.16	99.82	99.41	31.67	31.88	31.98
Pepper	99.29	99.11	99.67	31.44	31.24	31.67

6.4.2 Unified Average Change Intensity (UACI)

UACI is used to measure the average intensity difference in component and can be defined as

$$\text{UACI} = \frac{1}{N}\left[\sum_{i,j} \frac{\left|c(i,j) - c'(i,j)\right|}{2^L - 2}\right] \times 100\% \tag{6}$$

The proposed cryptosystem is analyzed with NPCR and UACI and the values are presented in Table 4.

From Table 4 shows clearly NPCR is achieved above 99% similarly UACI is achieved above 31%. So, the proposed encryption algorithm shows the extreme sensitivity when small changes in the original image and successfully resists the differential attack.

6.5 Correlation of Adjacent Pixels

For good cryptosystem, adjacent pixels in the encrypted image should be less correlated but in the original image, adjacent pixels are highly correlated. To test the correlation of adjacent pixels in the original image and cipher image, we randomly select 2000 pair's pixels in diagonal, vertical and horizontal direction. The following formula is used to calculate a correlation coefficient:

$$r_{xy} = \frac{\text{cov}(x,y)}{\sqrt{D(x)}\sqrt{D(y)}} \tag{7}$$

$$\text{where } \text{cov}(x,y) = \frac{1}{N}\sum_{i=1}^{N}(x_i - E(x))(y_i - E(y))$$

$$E(x) = \frac{1}{N}\sum_{i=1}^{N} x_i \quad D(x) = \frac{1}{N}\sum_{i=1}^{N}(x_i - E(x))^2$$

x and y are gray-level values of two adjacent pixels in the image.

Table 5 Correlation coefficients of adjacent pixels in the original Lena image and encrypted image

	Original image			Proposed method			Wei et al. [9]		
	Red	Green	Blue	Red	Green	Blue	Red	Green	Blue
H	0.9734	0.9631	0.9342	0.0018	0.0034	0.0072	0.0356	0.0763	0.0012
V	0.9621	0.9982	0.8912	−0.012	0.0034	0.0012	0.0127	0.0067	0.0098
D	0.9823	0.9102	0.9112	0.0046	−0.194	0.0041	0.0783	0.0562	0.0058

The results are presented in Table 5 and also we compared our results with Wei et al. [9].

From Table 5, the adjacent pixels in the original Lena image (Red, Green, and Blue channel) are highly correlated but in the encrypted image (Red, Green, and Blue channel), the adjacent pixels are less correlated and also better than [9]. So, our proposed encryption algorithm is successfully resisting the statistical attack.

7 Conclusion

In this paper, we have proposed simple color image encryption algorithm based on DNA operation and chaotic maps for Android environment. The Android environment is mainly for low processing devices like mobile phones, tablets, etc. So we have designed our proposed encryption algorithm to works in such environments. The main feature of the proposed algorithm is different encoding and decoding rules for every channel and diffusion operation, which was done between R, G, and B channels, not any values generated by chaotic equations. It reduces the execution time considerably. The experiment's results show the good stability against cryptographic attacks and also provide large key space. The proposed cryptosystem shows better results compared with Xiaopeng Wei et al. method. So it is suitable for any real-time applications.

References

1. Schneier B (1996) Applied cryptography-protocols, algorithms and source code, 2nd edn. Wiley, New York
2. Stallings W (2006) Cryptography and network security, 4th edn. Pearson Education
3. Matthews R (1989) On the derivation of a 'chaotic' encryption algorithm. Cryptologia 13(1):29–41
4. Li S, Chen G, Zheng X (2004) Chaos-based encryption for digital images and videos. In: Multimedia security handbook. CRC Press, LLC, Boca Raton, FL, USA, pp 133–67
5. Fridrich J (1998) Symmetric ciphers based on two-dimensional chaotic maps. Int J Bifurcat Chaos 8(6):1259–1284
6. Kanso A, Ghebleh M (2012) A novel image encryption algorithm based on a 3D chaotic map. Commun Nonlinear Sci Numer Simulat 17:2943–2959

7. Bigdeli N, Farid Y, Afshar K (2012) A novel image encryption/decryption scheme based on chaotic neural networks. Eng Appl Artif Intell 25:753–765
8. Liu H, Wang X (2011) Color image encryption using spatial bit-level permutation and high-dimension chaotic system. Opt Commun 284:3895–3903
9. Wei X, Guo L, Zhang Q, Zhang J, Lian S (2012) A novel color image encryption algorithm based on DNA sequence operation and hyper-chaotic system. J Syst Softw 85:290–299
10. Zhang Q, Guo L, Wei X (2010) Image encryption using DNA addition combining with chaotic maps. Math Comput Model 52:2028–2035
11. Zhang Q, Xue X, Wei X. A novel image encryption algorithm based on DNA subsequence operation. https://doi.org/10.1100/2012/286741
12. Zhang Q, Wei X (2013) A novel couple images encryption algorithm based on DNA subsequence operation and chaotic system. Optik 124:6276–6281

A Lucrative Sensor for Counting in the Limousine

R. Ajith Krishna, A. Ashwin and S. Malathi

Abstract In developing countries like India, the governments are in dire need of funds for all its projects. Hence it relies on a number of external factors such as exports, trading and FDI. On the other hand, the government also depends on internal factors such as tax from its citizens and indirectly from departments such as transportation. It is the responsibility of the government to provide transportation facilities like buses and trains to its citizens at a nominal charge and at the same time make a profit to run the government and its aided projects. To start from the lowest level, the sad truth is that the number of people travelling without tickets in buses is increasing at an alarming rate. This leads to a big loss for the transportation department and indirectly to the government which falls short of funds. When such losses are incurred due to defaulters in a large scale the government is forced to increase prices of other commodities to look at other sources of income from its citizens though hike in tax, interest rates, fuel price, VAT, etc. The proposed system reduces manpower by excluding checking inspector's role and their responsibilities. As we are all facing a lot of problems in identifying the number of passengers travelling without tickets the design of an automation system emerges with an in-built lucrative sensor that can count the number of passengers entering inside and leaving the limousine. Further, it can also detect the number of passengers who have taken tickets and who have not taken tickets, which will be handy to track down defaulter and make the bus ticketing system more efficient.

R. Ajith Krishna (✉) · A. Ashwin
Department of ECE, College of Engineering, Guindy, Chennai, India
e-mail: ajithkrishna1997@hotmail.com

A. Ashwin
e-mail: aashwin1@outlook.com

S. Malathi
Computer Science and Engineering, Panimalar Engineering College, Chennai, India
e-mail: malathi_raghu@hotmail.com

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_51

1 Introduction

In many applications like safety, transportation security, etc., embedded system plays a pivotal role as it is reliable system which can be isolated from hacking. By making subtle changes in the software, a general embedded system with a standard input and output configuration can be made to perform in a totally different manner. Embedded processors can be classified into two main categories namely, Ordinary microprocessors (μP) that use separate integrated circuits for memory and peripherals whereas Microcontrollers (μC) containing on-chip peripherals which gives benefits like low power consumption, size and cost.

The main objective is to design an efficient system wherein the number of persons coming in and out of a room or inside a vehicle is displayed on a screen. The presence of visitors is sensed using IR sensing mechanism and the entire counting mechanism is done by a microcontroller. This function is implemented using a pair of infrared sensors. LCD display placed outside the room displays this value of person count.

Nowadays, people counting system finds wide use for security purpose [1]. However, many algorithms have been implemented for the system design but clear approach is not yet followed in using these algorithms based on their drawbacks employed in people tracking and counting system. For the evaluation purpose, the number of people with digital camera makes it hard for automatic analysis of large images from many camera due to the high computing overhead. Moreover, video technology does not work in dark environments. In order to overcome this, people counting sensors [2, 3] are used in the proposed system. People counters provide valuable data related to the amount of traffic that flows through facilities. Placing people counters not only at entrances but also at separate displays and exhibits will give you more in-depth data about the popularity of each display.

In people counting system, the sensors are placed at the entry door. Whenever a passenger enters through the bus, there is a breakage in the infrared beam of the sensor [4]. By using this methodology, the total number of people inside the bus will be calculated and the total revenue of the trip will be estimated. Since the counting is done through the electronic equipment (sensor), there will be no manual errors [5] in total count unless there is a problem in the system. Thus, people counting system in transportation provide greater efficiency for counting the people.

2 Literature Survey

Estimating crowd density in an RF-based Dynamic Environment [6] was attempted by Yuan et al.; proposing a Device-free Crowd Counting approach without objects carrying any assistive device, which overcome the problem of high cost and low-light environments conducted by pattern recognition technologies on video surveillance. Accurate instantaneous and cumulative counts especially in crowded scenes [7] were determined using an integer programming method for estimating the instantaneous count of pedestrians crossing a line of interest in a video sequence. The video is first converted into a temporal slice image through a line sampling process. Then using regression function the number of people is estimated in a set of overlapping sliding windows on the temporal slice image which map from local feature to a count.

A concept-based indexing is a challenging approach that permits users to access videos that are conceptually related to the information provided in a search query to recognize the presence of concepts in a video segment for video retrieval over cross domains [8]. This is performed using a flexible energy optimization-based fusion and multicue fusion method integrates both the likelihood predicted by classifiers and high-order contextual-temporal relationships revealed from annotations and pseudo labels. Estimating the size of in homogenous crowd was the problem addressed by Chan and Vasconcelos [9] which are composed of pedestrians travel in different directions, without using explicit object segmentation or tracking is proposed.

An innovative approach called Occlusion Handling [10] is introduced to cross camera people for counting that can adapt itself to a new environment bypassing the manual inspection. In the analysis, a method for motion detection called ViBe as Universal Background Subtraction algorithm for video sequences was explained by Barnich et al. [11] that stores for each pixel, a set of values taken in the past at same location or in neighborhood. First, background subtraction [12] results estimated the number of people moving only slightly in a complicated scene. Second, an Expectation Maximization (EM) based method has been implemented to find the individuals in a low-resolution scene. The representation of each person in the scene is done using a novel cluster model. The current state-of-the-art image processing methods for automatic-behaviour-recognition techniques [13] is described by the system to monitor human activities in haulage. In the year 2009, Embedded Vision Modules for Tracking and Counting People showed the algorithm implementation [14] for a field-programmable gate array (FPGA)-based design for people counting using a low-level head-detection method.

In case of people in crowd areas, the crowd counting was found out using Bayesian Poisson Regression [15, 16] approach and it is analyzed by initiating a prior distribution on the weights of the linear function algorithm. Based on the flow velocity field estimation model and offline learning the algorithm has unified both of the LOI and ROI problems [17] together. Multi-object tracking which considers object detection and space-time trajectory estimation as a coupled optimization problem [18] were implemented by Leibe and Schindler. This proposed system employs a Minimum Description Length hypothesis selection framework, making the system to recover from mismatches and temporarily lost tracks. Segmentation and Tracking using multiple cameras in a crowded environment proposed [19] handled the inter-object occlusion problem.

Privacy-preserving crowd monitoring [20] handled the in homogenous crowds in different directions and a Gaussian process (GP) is used to regress feature vectors to the number of people per segment. High accuracies obtained by using motion capture and surveillance which automatically track the articulations of people from a video sequence [21]. This creates ambiguity while examining the people in both frame and in estimation of their layout. Bayesian clustering algorithm [22] assessing the value of motion and the information content of unsupervised data-driven has been left out as the control. In the survey, an efficient technique for Pedestrian detection [23] by Leibe et al.; specifically address the task of detection in crowded scenes. It handles the problem of detecting people in crowd with severe overlaps. Stauffer and Grimson focused on motion tracking and showed how one can use observed motion to learn patterns of activity [24] in a site. In enclosed surroundings this system is used to track people alone, whereas both people and vehicles are tracked in open atmosphere. Also, the system is used even for finding the fish in a tank, ants on a floor, and remote control vehicles in a lab setting.

3 Proposed Work

In recent days, people counting system is being generally used for security purpose. Many algorithms have been designed and implemented, but still there lies a confusion in selecting the suitable algorithms considering the advantages and disadvantages. Moreover, video technology does not work in dark environments. In order to overcome this, people counting sensors are used in the proposed system.

Figure 1 shows the architecture diagram for the system Where the IR sensor is given as input to the port1 of the ATMEL microcontroller. When the conductor is issuing tickets to passenger, count value of issued tickets will be stored in the database at each bus stop. The system detects the frauds that are travelling without tickets by comparing the difference between count value of passengers inside the bus and count value of issued tickets at each stop. The components included in the system are ATMEL microcontroller represented as 'AT89S52' which is a typical 8051 microcontroller. The IR sensor is included in the system that produces or detects infrared radiation in order to sense some aspect of its environment.

ALGORITHM: COMMUTER SENSOR (p1, p2. . . pS). Given a Set of N Training Examples with their S types of Distance Descriptors.

```
Step 1: Initialize Ptc=0
Step 2: Initialize s=0
Step 3: while(s==x)
//varying the distance by adjusting the probe
Step 4: for i=1 to P
Step 5: if (! ir_in) //infrared beam interrupted at the entrance
Step6:Ptc++
Step 7: else if (! ir_out)
//infrared beam interrupted at the exit
Step 8: Ptc- -
Step 9: else Print Ptc
Step 10: end if
Step 11: end for
Step 12: end while
Step 13: return Pi count value
```

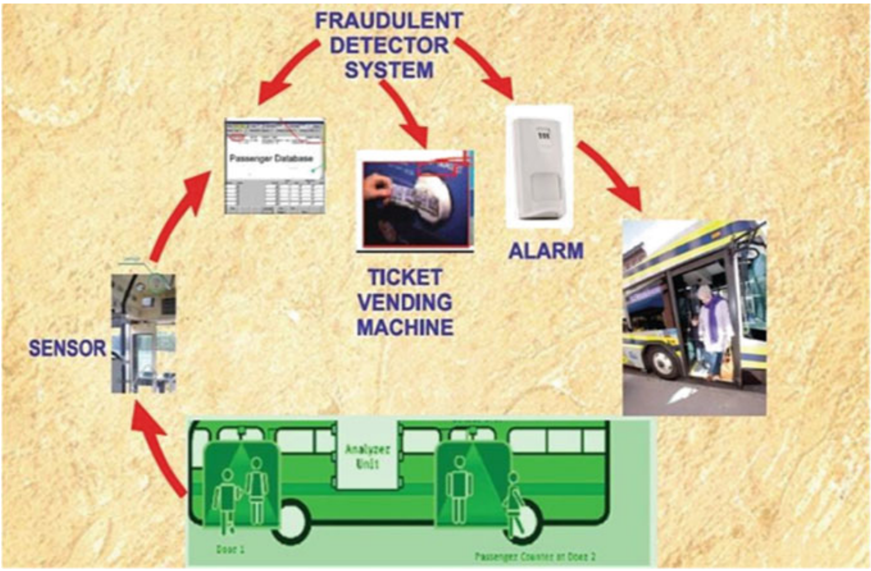


Fig. 1 Architecture framework for counting in the Limousine

A set (p1, p2, p3 … pS) of N training examples is specified as an input to the sensor with their types of distance descriptors, where p represents the number of people entered inside the bus. This parameter is considered for N number of samples and gets repeated until the limousine reaches its destination and Δs is the symbol of distance that can be attuned by varying their ranges.

ALGORITHM: TRAVELLING WITHOUT TICKETS combines (p1,p2, . . ., pS) and (t1,t2, . . . ,tD).Given a Set of N Training examples with their S types of distance descriptors.

```
Step 1: Initialize place=source
Step 2: While (! place=destination)
Step 3: for i=1 to 5
        //iterate the loop for every 5 minutes
Step 4: for j=1 to T
        //loop iterates for generation of each ticket
Step 5:         if (! BS)
        //check whether bus stop has not arrived
Step 6:         wt=Ptc - tc
        //compute the travelers who haven't taken tickets
Step 7: end if
Step 8: end for
Step 9: end for
Step10: end while
Step11: return total wt value
```

A set (p1, p2, p3 … pS) of N training examples is specified as an input to the sensor with their types of distance descriptors, where p represents the number of people entered inside the bus. This parameter is considered for N number of samples and gets repeated until the limousine reaches its destination and s is the symbol of distance that can be attuned by varying their ranges. Initially, a parameter named Ptc represents the people total count to be zero then validates the while condition loop by checking whether s equals the varying distance x. The count of without tickets can be calculated for every stop till the destination place.

4 Results and Discussion

Labview finds its application in science and engineering field for creating custom applications which interact with real-world data or signals. Instead of sequential line by line operation, G is an efficient graphical dataflow language in which nodes (operations or functions) operate on data once it becomes available. Table 1 represents the estimation of number of passengers travelling without tickets is detected.

Figure 2 represents the plotted graph used for estimating the number of passengers travelling without tickets by indicating the blue bar for number of passengers entering inside the bus and red bar for the tickets issued for passenger entered inside the limousine.

Figures 3 and 4 depict the outcome of the passengers travelling from source station and final destination where the passengers without tickets are verified.

Table 1 Number of passengers without tickets

Bus stop	Time (min)	Iteration	
		No. of passengers	No. of tickets issued
s1	3	10	8
s2	2	8	6
s3	2.5	9	7

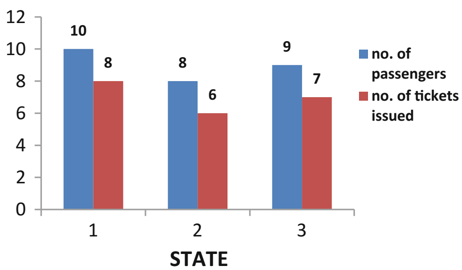


Fig. 2 Passengers without tickets based on time

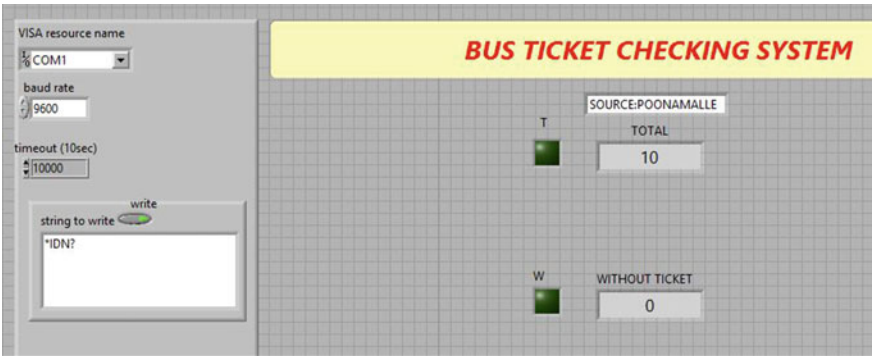


Fig. 3 Passengers at source station

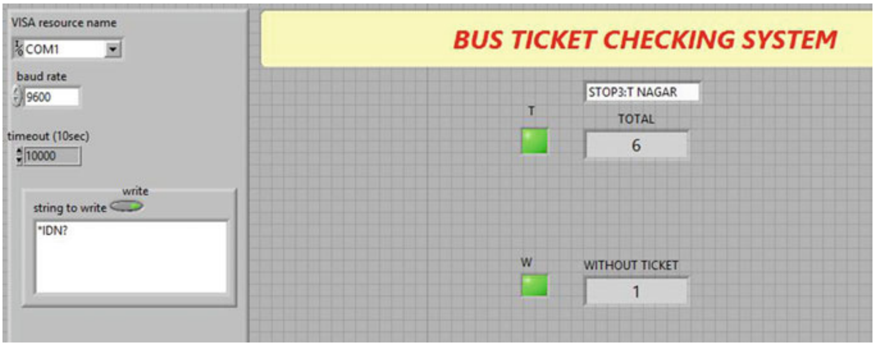


Fig. 4 Passenger without tickets at the destination

5 Conclusion

Transportation department is facing numerous problems in identifying the number of passengers travelling without tickets, an automation system has been designed with an in-built commuter computational sensor for counting number of passengers entering and leaving and detecting number of passengers travelling with tickets and without tickets which reduces manpower by excluding checking inspector's role and their responsibilities and loss of transportation's fund. There is still much to improve the system in identifying passengers who are travelling without tickets by using a fixed camera inside that can work by integrating with ATMEL microcontroller combining the facial recognition techniques in image processing it will also feature a notification system called Rap Back, which will give investigators live updates of any given criminal's movements.

In a bid to modernize US detective work, an alternate replacement for a countrywide fingerprint database using the current system. This recognition can also be implemented by scanning the social security card like Adhaar card that contains details like in an individual file followed by linking them to personal and biographic data like name, home address, ID number, immigration status, age, race, etc. The other federal agencies also share the database which contains new records that reveals that this is capable of processing 55,000 direct photo enrollments and also conducting numerous searches like tens of thousands every day.

References

1. Leibe B, Schindler K, Van Gool L (2007) Coupled detection and trajectory estimation for multi-object tracking. In: Proceedings of IEEE international conference computer vision, 2007, pp 1–8
2. Yang DB, González-Bãnos HH, Guibas LJ (2003) Counting people in crowds with a real-time network of simple image sensors. IEEE Conf Comput Vis Pattern Recogn 2
3. Li Y, Huang C, Nevatia R (2009) Learning to associate: hybridboosted multi-target tracker for crowded scene. In: Proceedings of IEEE international conference
4. Liu X, Taommingli Song D, Zhang L, Bu J, Chen C (2014) Learning to track multiple targets. IEEE Trans Neural Netw 20
5. Muñoz-Salinas R, Medina-Carnicer R, Madrid-Cuevas FJ, Carmona-Poyato A (2009) Multi-camera people tracking using evidential filters. Elsevier Inc.
6. Yuan Y, Zhao J, Qiu C, Xi W (2013) Estimating crowd density in an RF-based dynamic environment. IEEE Sens J 13(10)
7. Ma Z, Chan AB (2013) Crossing the line: crowd counting by integer programming with local features. In: Proceedings of conference computer vision pattern recognition, pp 2539–2546
8. Weng M-F, Chuang Y-Y (2012) Cross-domain multicue fusion for concept-based video indexing. IEEE Trans Pattern Anal Mach Intell 34(10):1927–1941
9. Chan AB, Vasconcelos N (2012) Counting people with low-level features and Bayesian regression. IEEE Trans Image Process 21(4):2160–2177
10. Lin T-Y, Lin Y-Y, Weng M-F, Wang Y-C, Hsu Y-F, Liao H-YM (2011) Cross camera people counting with perspective estimation and occlusion handling. In: Proceedings of IEEE international workshop information forensics security, pp 1–6, Nov/Dec 2011
11. Barnich O, Van Droogenbroeck M (2011) vibe: a universal background subtraction algorithm for video sequences. IEEE Trans Image Process 20(6):1709–1724
12. Hou Y-L, Pang GKH (2011) People counting and human detection in a challenging situation. IEEE Trans Syst Man Cybern A Syst Humans 41(1):24–33
13. Candamo J, Shreve M, Goldgof DB, Sapper DB, Kasturi R (2010) Understanding transit scenes: a survey on human behavior-recognition algorithms. IEEE Trans Intell Transport Syst 11(1)
14. Vicente AG, Muñoz IB, Molina PJ, Galilea JLL (2009) Embedded vision modules for tracking and counting people. IEEE Trans Instrument 58(9)
15. Chan AB, Vasconcelos N (2009) Bayesian Poisson regression for crowd counting. In: Proceedings of IEEE international conference computer vision 545–551
16. Zhao T, Nevatia R (2003) Bayesian human segmentation in crowded situations. Proc IEEE Conf Comput Vis Pattern Recogn 2:459–466
17. Cong Y, Gong H, Zhu S-C, Tang Y (2009) Flow mosaicking: realtime pedestrian counting without scene-specific learning. In: Proceedings of IEEE conference computer vision pattern recognition, June 2009, pp 1093–1100
18. Leibe B, Schindler K, Cornelis N, Van Gool L (2008) Coupled object detection and tracking from static cameras and moving vehicles. IEEE Trans Pattern Anal Mach Intell 30(10)
19. Zhao T, Nevatia R, Wu B (2008) Segmentation and tracking of multiple humans in crowded environments. IEEE Trans Pattern Anal Mach Intell 30(7):1198–1211

20. Chan AB, Liang ZSJ, Vasconcelos N (2008) Privacy preserving crowd monitoring: counting people without people models or tracking. Proc IEEE Conf Comput Vis Pattern Recogn 1–7
21. Ramanan D, Forsyth DA, Zisserman A (2007) Tracking people by learning their appearance. IEEE Trans Pattern Anal Mach Intell 29(1):65–81
22. Brostow GJ, Cipolla R (2006) Unsupervised Bayesian detection of independent motion in crowds. Proc IEEE Conf Comput Vis Pattern Recognit 1:594–601
23. Leibe B, Seemann E, Schiele B (2005) Pedestrian detection in crowded Scenes. Proc Conf Comput Vis Pattern Recognit 878–885
24. Stauffer C, Grimson E (2000) Learning patterns of activity using real-time tracking. IEEE Trans Pattern Anal Mach Intell 22(8):747–757

Energy Expenditure Calculation with Physical Activity Recognition Using Genetic Algorithms

Y. Anand and P. P. Joby

Abstract Physical health is associated with physical activity, physical activity also ensures the wellbeing of the humans, physical activity is recognized using body worn sensors, and three Inertial Measurement units (IMU) are used to capture the data from the sensors. The activity recognition chain consists of Data Acquisition, Preprocessing, segmentation, Feature extraction, and Classification. Different levels of research are carried out on each stage. In feature selection genetic algorithms are used but the paper proposing the memetic algorithms an enhanced version of the genetic algorithms with local search in the each stage of genetic algorithm. This technique shall eliminate the chances of energy loss and consequently increase efficiency of the current system.

1 Introduction

The physical activity is solely related to the human health, but unfortunately advancements in transportation and technology made everything available at doorsteps and fingertips, reducing physical activity thereby reducing physical health. Several works were done in the fields of relating physical activity and physical health, works of Janssen and LeBlanc [1] stating that even physical activity can improve the health considerably. Next big challenge is to monitor and record the physical activity, so we could have qualitative analysis of the physical activity, this can be done using three IMUs (Inertial measurement units). Specialized systems can record these activities and can be used for the further processing, machine learning data set is from the available dataset is done using running data through Activity Recognition Chain

Y. Anand (✉)
Computer Science and Engineering, Mar Baselios Christian College of Engineering and Technology, Peermade, India
e-mail: michaelanand2020@gmail.com

P. P. Joby
Mar Baselios Christian College of Engineering and Technology, Peermade, India
e-mail: jobymone@gmail.com

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_52

(ARC) [2]. Energy expenditure will be different for different activities based on MET value. Calculating accurate MET value with minimal use of sensors and is the idea explained in the paper.

2 Overview of Physical Activity and Health Relation

The importance of keeping physical health is much important for well-being and keeping life expectancy high, reduced physical activity in any age group will lead to the serious health issues. People are aware of it but lack motivation, giving them a quantitative detail with the aid of technology will aid in improving activity and increasing health. Different physical activity requires different level of energy and some intense activities requires only small time to burn more calories (energy spent in a activity) and light activities requires more time to do the same. Building a specific set of activities to burn more calories in a shorter time is the key aspect. Doing the activities for burning more energy will increase the health, thereby reduce the risk of many dangerous lifestyle disorders like cholesterol, blood pressure, etc. In order to find activities with more energy burning relation we need to know some value to calculate the energy. The collected data from the sensors will be converted into some useful features, which can be analyzed in an efficient manner.

3 Physical Activity and Energy Expenditure

MET or metabolic equivalent is amount of energy we spent in each activity, as per m. Jette et al. [3] work about MET, one MET is equal to amount of oxygen intake when our body being rested, taking this value as unity other activities MET can be calculated as the multiple of these, one MET means the 3.5 ml of O_2 (oxygen) one MET is equivalent to 17.5 W (watts) and one Watt is equal to 0.01435 kcal of energy Compendium of Physical Activities has provided the provided MET values [4] and Kozey et al. proposed work of calculating MET Value from the accelerometer outputs. Getting the Data from the sensors and processing it to find the values.

4 Physical Activity Recognition Using Sensors

The human physical activity recognition been in interest of the researchers in the early 20s, later many research has been done on the physical activity recognition using sensors, Some devices from gaming consoles are very well known for tracking and analyzing the human activity and also found some space in fitness applications, devices like these enhanced the computer human interaction. Main characteristics of human activity recognition systems are given in Bulling et al. [5].

4.1 The Activity Recognition Chain

The ARC (activity recognition chain) is the sequence of processing the signal, specific to machine learning techniques the activity recognition chain explains how the raw data from the sensors are converted to the useful data, Fig. 1 explains how dataset is generated with the help of the sensors.

Sensor Data Acquisition and Preprocessing The main objective is to capture the data from the sensors that are worn on subject's body, three colibri wireless sensors were used on this process, position of the IMU (inertial measurement units) [6].

- 1 IMU over the wrist on the dominant arm
- 1 IMU on the chest
- 1 IMU on the dominant side's ankle.

The sensors should have components like accelerometer, gyroscope, temperature sensor, magnetometer, etc., All these components were collected with the corresponding attributes from the body, 9 (8 men and 1 women aged 27.22 ± 3.31) subjects are requested to do certain activities.

Each IMU collects 17 columns of data in total which consists of

- 1 temperature (°C)
- 2–4 3D-acceleration data (ms^{-2}),
- 5–7 3D-acceleration data (ms^{-2})
- 8–10 3D-gyroscope data (rad/s)
- 11–13 3D-magnetometer data (μT)
- 14–17 orientation (invalid in this data collection).

When building a dataset that can be useful for the later processing we should also add some useful attributes like timestamp, activity id for uniquely identifying each activity.

Activity IDs:

- 1 lying
- 2 sitting
- 3 standing

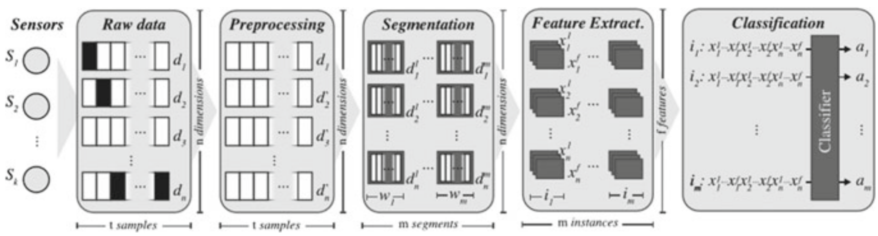


Fig. 1 Activity recognition chain

- 4 walking
- 5 running
- 6 cycling
- 7 Nordic walking
- 9 watching TV
- 10 computer work
- 11 car driving
- 12 ascending stairs
- 13 descending stairs
- 16 vacuum cleaning
- 17 ironing
- 18 folding laundry
- 19 house cleaning
- 20 playing soccer
- 24 rope jumping
- 0 other (transient activities).

HR (Heart Rate) is also included so that the bpm (beats per minutes) will be recorded by the heart rate monitor. All these data are synchronized and labeled data from all the sensors are merged into a single file. finally there will be 54 columns of data that contains the following data:

- 1 timestamp (s)
- 2 activity ID
- 3 heart rate (bpm)
- 4–20 IMU hand
- 21–37 IMU chest
- 38–54 IMU ankle.

Data Segmentation The data need to be segmented in order to processing, data collection is continuous and repeated over 0.01 s (seconds) so we have to specify a start time t_1 and end time t_2 and all the data collected from this time should be considered as a segment [5]. The whole data collected can be converted to some useful segments, a single segment can be defined as,

$$W_i = (t_1, t_2) \tag{1}$$

W is set of segments were

$$W = (w_1 \ldots w_m) \tag{2}$$

Segmenting the data can be done based on different methods, we use sliding window method. The size of sliding window is fixed as 512 (5.12 s), that is the sensor data of 5.12 s are grouped together.

Feature Extraction This stage consist of converting the segments into some useful features, features are brief abstract of the raw data, data collected can be represented as features like mean and variance, running the DFT (Discrete Fourier Transform) over the segmented data of every signal and dimension using the (FFT) Fast Fourier Transform [6] in the feature selection challenges arise due to various factors selecting the correct feature selection techniques this is directly linked to the accuracy of the whole data processing.

Classification This is the final process in ARC. In this data should be classified for each person and this data can be used for cross person classification done based on the feature extraction. A technique known as extra trees in machine learning library is used for the work Alejandro et al. [6] and obtained an accuracy of 97.45 in the final result.

5 Calculating the Energy Expenditure

To find the energy spend on each activity in terms of calories or joules, this could be very useful for the user to assess the performance every day. Calculating the energy expenditure is given as

$$\text{Total amount of calorie burned} = \text{MET} * \text{Body weight} * \text{time duration} \quad (3)$$

Oxygen intake is directly related to the energy expenditure, for calculating MET, accelerometer data and heart rates can be used [4]. With the help of accelerometer data and wireless unit that measures breath-by-breath gas exchange. MET value can be calculated for each individual. This Data can be compared to user's previous data or other persons with cross person validation. More energy spent on the activity if it consumes more oxygen, the oxygen gas flow is key factor to compute the total energy expenditure.

6 Conclusion and Future Work

The physical activity is essential to maintain the physical health, new age of technology will aid us in tracking and improving our health by analyzing our physical activity, the different techniques used for acquisition and processing of data and how this data can be used to calculate the energy is explained. Future work on this paper would be by building a machine learning system can be used for learning the user movements and an artificial intelligence system can be used for the giving continuous suggestion to the user about the his/her physical activity and it can be programmed such a way that we could do a intense work in small amount of time.

References

1. Janssen I, LeBlanc AG (2010) Systematic review of the health benefits of physical activity and fitness in school-aged children and youth. Int J Behav Nutr Phys 7(1):40
2. Reiss A, Stricker D (2012) Introducing a new benchmarked dataset for activity monitoring. In: Proceeding of ISWC'12, pp 108–109
3. Jette M, Sidney K, Blumchen G (1990) Metabolic equivalents (METS) in exercise testing, exercise prescription, and evaluation of functional capacity. Clin Cardiol 13(8):555–565
4. Kozey SL, Lyden K et al (2010) Accelerometer output and MET values of common physical activities. US National Library of Medicine National Institutes of Health. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2924952/
5. Bulling A, Blanke U, Schiele B (2014) A tutorial on human activity recognition using body-worn inertial sensors. ACM Comput Surv 46(3):33:1–33:33
6. Baldominos A, Isasi P, Saez Y (2017) Feature selection for physical activity recognition using genetic algorithms. IEE Explore

Improved Intrinsic Image Decomposition Technique for Image Contrast Enhancement Using Back Propagation Algorithm

Harneet Kour and Harpreet Kaur

Abstract The technique of intrinsic image decomposition is based on the illumination value of the image. The histogram equalization value of the input image is calculated to increase the image contrast. In this research work, the back propagation algorithm is applied for the calculation of histogram equalization. The iterative process of back propagation is executed until error is reduced for the histogram equalization calculation. The simulation of the proposed modal is performed in MATLAB. The performance of proposed modal is compared in terms of PSNR and MSE.

1 Introduction

A mechanism in which an image is converted into digital form is known as image processing. In order to enhance the image or to extract important information present within, various operations are performed on that image. In order to provide certain benefits within applications, image processing has been growing in demand. Digital image processing techniques are now applicable within numerous other applications along with medicine and space programs [1]. For easy interpretation of X-rays and various other images available from different sources, various computer procedures are applied which help in enhancing the contrast or other features present within the image. Within most of the image processing applications, image enhancement is the most interesting area of research. The details are made more obvious and specific features that are of interest are highlighted with the help of applying image processing techniques [2]. The images that are ambiguous and uncertain can include different aspects within them. The qualities of an image are enhanced with respect to contrast, brightness properties, removal of noise and so on, by applying image

H. Kour
ECE Department, Chandigarh University, Gharuan, India

H. Kaur (✉)
CSE Department, Chandigarh University, Gharuan, India
e-mail: harpreet8307@gmail.com

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_53

enhancement operation. An image might be blurred in some cases due to which the objects present within it might not be visible clearly. Thus, there is a need to enhance that image such that its contents can be visible. In order to provide solutions to this issue, contrast enhancement is one important research aspect [3]. There are several contrast enhancement techniques used among which some are explained below:

a. Histogram Equalization (HE): In order to enhance the contrast of an image, the most popular technique known is HE. On the basis of probability distribution of input gray levels, the available fray levels are mapped within this technique. The dynamic range of image's histogram is flattened and stretched through this technique due to which the complete contrast enhancement is achieved [4]. The image is treated globally through the HE technique which is a major advantage. The images that include within them both bright or both dark backgrounds can attain better results when this technique is used.
b. Contrast-Limited Adaptive Histogram Equalization (CLAHE): An adaptive contrast enhancement technique which is based on the adaptive histogram equalization which is enhancement to the traditional HE technique is known as CLACHE. Instead of computing the complete image, a specific section of an image which is known as tile is computed using various histograms in this technique. In order to redistribute the pixel values of an image, the contrast of each tile is improved. Further, for removing any artificially added boundaries, the bilinear interpolation is utilized in order to combine the neighboring tiles [5]. In order to prevent noise amplification that is present within an image, the contrast is ensured to be limited within the homogeneous regions. The local contrast of an image is thus enhanced and more specific details are achieved through this technique. Instead of focusing on the overall contrast of an image, the local contrast is only focused on through this method.
c. Morphological Enhancement: A new technique is generated in order to solve the various issues arising in this field by applying mathematical morphology to the image processing and analysis methods. On the basis of set theoretic concepts of shape, this approach is generated [6]. The objects that are present in an image are considered to be sets as per the morphology technique. The machine vision and recognition processes can easily use mathematic morphology since the objects and the features present in them can be identified on the basis of their shapes.

2 Literature Review

Wang et al. (2017) proposed an enhanced to the image contrast enhancement preprocessing approach such that the model parameters can be derived in automatic manner from the cell video sequences as per each frame [7]. This technique is derived in order to avoid issues that arise due to low image intensity contrast characteristics within cell images. Two image sequences were tested and comparisons were made against Viterbi-based segmentation and the proposed technique. As per the evaluations it was seen that there was around 37 and 33% of enhancement in accuracy through application of proposed technique.

Murinto et al. (2017) proposed a novel approach in order to enhance the leather image. On the basis of un-sharp masking algorithm, piecewise linear contrast stretch technique is utilized here [8]. Upon four different animal leather images of animals namely lizard, crocodile, cow and goat, the experiments are conducted here. Comparisons are made amongst other existing enhancement techniques and it is seen as per the simulation results that the proposed technique is better in comparison to proposed technique in terms of PSNR, MSE and SSIM values.

Nnolim (2017) proposed an enhanced algorithm for single underwater image enhancement. More generalized global contrast enhancement is utilized in the PDE-based framework in order to develop adaptive as well as completely automated algorithms [9]. The color correction, edge as well as contrast enhancement results are improved as per the results generated by applying proposed technique. The underwater images that include flat color histograms are very difficult to improve by utilizing traditional techniques can be enhanced by applying this proposed technique.

Fornes et al. (2013) proposed a novel technique in order to enhance and eliminate the show-through scheme present within gray-scale documents [10]. Only one side of the page is utilized by this method and there is no need to eliminate the framing borders. There is a need to tune only one parameter within this approach. As per the simulation results it is seen that the readability of documents is enhanced due to which the unreadable words can be recovered. Also, the ambiguities of the document can be solved and the amount of time that is required to read each document is also minimized by using this proposed technique.

Rosenberger (2012) proposed the utilization of Spartan 6 FGPA within the developed camera. In order to attain higher processing power, the properties of DSP 48 slices were utilized [11]. The frame rate of CCD sensor is the speed limitation for this application. Lots of additional sampling time is saved with the help of generating VCE image. A low uncertainty of optical measurement is achieved when a well illuminated image is required along with various sub-images. The results achieved show that this mechanism helps in eliminating previous problems and providing enhanced results.

In Fig. 1, the framework of proposed contrast enhancement method is shown. Initially, an input color image I is converted into HSV representation. Further, the proposed intrinsic decomposition model is used that decomposes the value (V) channel image into illumination (L) and reflectance (R) layers as secondary step. An adjusted L layer which is represented as L_a is generated when L layer is adjusted using Gamma mapping function which is the third step. In order to generate the enhanced V channel image which is represented here as V_e, the adjusted L_a is multiplied by reflectance layer R. In order to enhance the local contrast of V_e the contrast-limited adaptive histogram equalization (CLAHE) [12] is utilized due to the fact that the mapping function can be performed globally. $\hat{V}_e$ is used to denote the enhanced result. The technique of back propagation will be applied to calculate the histogram equalization. The technique of back propagation take input the pixel number of pixel value

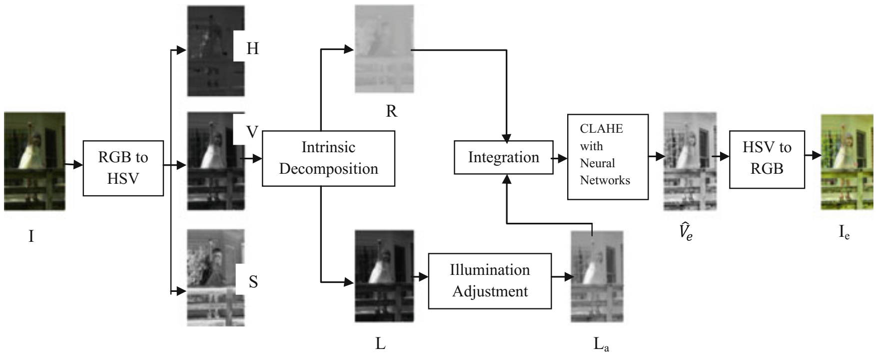


Fig. 1 The framework of the proposed contrast enhancement method

to generate the final equalization value. The formula is given in Eq. (1) to calculate actual value of equalization

$$\text{Actual equalization value} = \sum_{\substack{x=0\\w=0}}^{\substack{x=n\\w=n}} x_n w_n + \text{bias} \tag{1}$$

To calculate the histogram equalization, we need to define the desired value. The final equalization will be at which error is least. The error will be calculated with the Eq. (2)

$$\text{Error} = \text{Desired Equalization value} - \text{Actual equalization value} \tag{2}$$

At the end, the final result I_e is generated by transforming the enhanced HSV image into RGB space. Intrinsic decomposition and illumination adjustment are the two major modules of this proposed mechanism which are explained further.

a. Intrinsic Image Decomposition

In order to compute the equation $V = L.R$, one image is decomposed into reflectance and illumination layers within the Intrinsic image decomposition. In the form of constraints that involve the neighboring pixels with same color to have similar reflectance and the neighboring pixels that have similar illumination, intrinsic decomposition is proposed here [13]. Thus, the intrinsic decomposition issue is formulated here within Eq. (3), as a minimization problem of the energy function. Here, vector form of V, L, and R are used which are v, l, and r.

$$\begin{aligned} &\min_{l,r}\ E(l,r) = E_r(r) + \mu E_l(l) + \theta E_d(v; l, r) + \beta E_\circ(l, l_\circ), \\ &\text{s.t.} \qquad 0 \le r \le 1, \end{aligned} \tag{3}$$

Here, the weighting parameters are denoted by $\mu, \theta,$ and β. Upon the reflectance and illumination layers, μ and θ are regularizer. In order to ensure the reliability of decomposition, β terms is used. l_2 norm penalty which is $\left(\|v - l\ .\ r\|_2^2\right)$ is utilized instead of previous equality constraint which was $v = l\ .\ r$ within the intrinsic decomposition such that the noise can be tolerated. The noise might decompose otherwise. In order to constrain the scale of l, the final term $E_\circ = \left(\|l - l_\circ\|_2^2\right)$ is used. The chromatic normalization value which is $\sqrt{I_r^2 + I_g^2 + I_b^2}$ is set for $l_\circ$ here.

The reflectance later is constrained to be piecewise constant as per the similarity of color by the $E_r(r)$ term. For instance, at pixel i, the reflectance value is denoted by r_i. The representation of this constraint is shown in Eq. (4).

$$E_r(r) = \sum_i \sum_{j \in N(i)} w_{ij} \left\| r_i - r_j \right\| 1 \tag{4}$$

Here, for pixel i, the neighborhood is $N(i)$. For pixel i and j, the similarity of chromatic value is measured by w_{ij}. The weight values are increased here such that the different amongst r_i and r_j can be penalized for the pixels that have similar colors. Thus, the weighting function is defined in Eq. (5) as:

$$w_{ij} = \exp\left(-\frac{\|f_i - f_j\|_2^2}{2\sigma^2}\right) \tag{5}$$

Here, the value of pixel i is represented by f_i which can be denoted as $f_i = [\tau l_i, a_i, b_i]^{\mathrm{T}}$. The values of τ and σ are constant. It is to be ensured that $\tau < 1$ in order to minimize the impact of illumination variations on the color similarity measurement.

The isotropic total variation is utilized in order to enforce that the illumination is smooth with the application of $E_l(l)$, and is defined by Eq. (6) as:

$$E_l(l) = \|D_x l\|_2^2 + \left\|D_y l\right\|_2^2 \tag{6}$$

Here, along the horizontal and vertical directions, D_x and D_y represent the matrix representation of derivative operators respectively.

b. Illumination Adjustment

In order to enhance the image details, the adjustment of illumination values is the next major task to be performed after calculating the reflectance and illumination layers. Through the results obtained it is seen that the by preserving the lightness order, all the dark areas are darkened. Within the bright areas however, the intensities are compressed such that there is minimization of variation of large intensities. Thus, as a result, within the bright areas, the details will be lost mainly within the Log and

Fig. 2 **a** Input image. **b** Intrinsic image decomposition. **c** Improved intrinsic image decomposition

Sigmoid functions [13]. Thus, Gamma function is utilized here for adjusting the illuminations which is represented in Eq. (7).

$$L_a = 255 \times (L/255)^{1/\gamma} \tag{7}$$

Here, for conducting experiments, the value of γ is set as 2.2.

3 Experimental Results and Analysis

The performance of proposed technique is evaluated in this section. The performance of proposed technique will be compared with the intrinsic image decomposition technique. The performance of the proposed modal can be viewed in both subjective and objective form. The proposed algorithm can be tested on the dark image for the contrast improvement.

As shown in Fig. 2a is the input image which is the dark image whose contrast needs to improve. The technique of intrinsic image decomposition is applied in Fig. 2b. The Improved Intrinsic image decomposition technique is applied in Fig. 2c which gives best result as compared to existing technique.

As shown in Fig. 3, the Intrinsic Image decomposition technique is compared with the proposed intrinsic image decomposition technique in terms of PSNR and MSE. The simulation results illustrated that proposed technique performs well in terms of all parameters.

4 Conclusion

In this paper, improvement in the intrinsic image decomposition technique is proposed. The intrinsic image decomposition technique is based on the illumination layer in which histogram equalization is calculated to image contrast. In this research work, the improvement in the intrinsic image decomposition technique is proposed

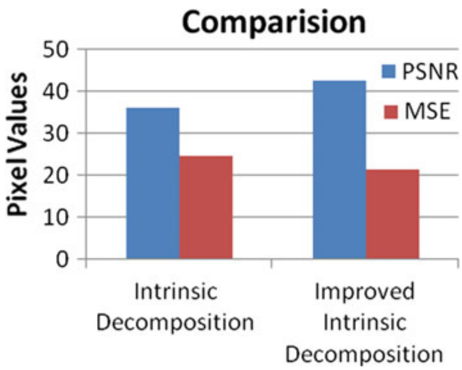


Fig. 3 Comparison of techniques

based on back propagation technique. The back propagation technique will be executed in the iterative manner until error gets reduced for the calculation of histogram equalization. The iteration at which error is least is defined as the final histogram equalization point. The simulation results shows that proposed technique performs well in terms of PSNR and MSE than existing intrinsic technique.

References

1. Cheng P, Cui A, Yang Y, Luo Y, Sun W (2017) Recognition and classification of coating film defects on automobile body based on image processing. In: 2017 10th International congress on image and signal processing, biomedical engineering and informatics (CISP-BMEI 2017)
2. Purohit AD, Khandare ST (2017) A survey on different color image segmentation techniques using multilevel thresholding. IJCSMC 6(4):267–273
3. Ritika, Kaur S (2013) Contrast enhancement techniques for images–a visual analysis. Int J Comput Appl 64(17)
4. Liang Z, Liu W, Yao R (2016) Contrast enhancement by nonlinear diffusion filtering. IEEE Trans Image Process 25(2):673–686
5. Li Y, Guo F, Tan RT, Brown MS (2014) A contrast enhancement framework with JPEG artifacts suppression. In: European conference on computer vision. Springer, pp 174–188
6. Yun S-H, Kim JH, Kim S (2011) Contrast enhancement using a weighted histogram equalization. In: 2011 IEEE international conference on consumer electronics (ICCE). IEEE, pp 203–204
7. Wang X, Cheng E, Burnett IS (2017) Improved cell segmentation with adaptive Bi-Gaussian mixture models for image contrast enhancement pre-processing. IEEE
8. Murinto, Winiarti S, Ismi DP, Prahara A (2017) Image enhancement using piecewise linear contrast stretch methods based on unsharp masking algorithms for leather image processing. In: 2017 3rd International conference on science in information technology (ICSITech)
9. Nnolim UA (2017) Improved partial differential equation-based enhancement for underwater images using local–global contrast operators and fuzzy homomorphic processes. IET Image Process 11(11):1059–1067
10. Fornes A, Otazu X, Llados J (2013) Show-through cancellation and image enhancement by multiresolution contrast processing. In: 2013 12th International conference on document analysis and recognition

11. Rosenberger M (2012) Virtual contrast enhancement intelligent illumination adjustment processing with field programmable gate array based camera systems for imaging applications enhancing contrast in multi AOI applications. IEEE
12. Reza AM (2004) Realization of the contrast limited adaptive histogram equalization (CLAHE) for real-time image enhancement. J VLSI Sig Process Syst Sig Image Video Technol 38(1):35–44
13. Yue H, Yang J, Sun X, Wu F, Hou C (2016) Contrast enhancement based on intrinsic image decomposition. IEEE Trans Image Process

Low-Power High-Speed Hybrid Multiplier Architectures for Image Processing Applications

U. Saravanakumar, P. Suresh and V. Karthikeyan

Abstract Multipliers play an imperative role in communication, signal and image processing, and embedded ASICs. Generally, multipliers are designed through various steps and they are occupying more area in the hardware, consumes more power, and causes an effect on performance. This paper is aimed to implement low-area and high-speed multiplier using various data compressors in partial product stages and tested for image processing. To suppress the vertical dimension of the partial product stage in multiplier, Sklansky adder is considered for the last stage and five hybrid multiplier architectures (HyMUL1–HyMUL5) have been implemented. For application verification, the two grayscale images are given as the inputs of the proposed multipliers and produce a new image which is an overlap of the two input images. The comparative analysis indicates that the proposed multiplier HyMUL2 consumed less area compared to other multipliers and its speed is also improved. The obtained new overlapped image using proposed multiplier HyMUL2 has high PSNR and low NMED.

1 Introduction

Most of the computing processes for real-time applications are implemented using digital building blocks such as adders, multiplier, comparators, etc., thus operating with a high degree of consistency and exactness. However, many applications including multimedia can accept a small number of errors and inaccuracy in computation but produce valid results. Since multimedia applications work with inaccurate mod-

U. Saravanakumar (✉) · P. Suresh · V. Karthikeyan
Department of ECE, Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Chennai 600062, India
e-mail: saran.usk@gmail.com

P. Suresh
e-mail: suresh3982@gmail.com

V. Karthikeyan
e-mail: vkarthikeyan652@gmail.com

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_54

els, designers have approached several architectures for approximate computation to achieve the simplest design process and other hardware-related benefits like power consumption and area occupancy [1]. The key element in digital building blocks is data compressors and they have been studied in detail for approximate implementation [2]. Since the multipliers for handheld devices require more number of adders to minimize the height of partial product, the demand for low area multiplier is continuously increasing.

A multiplication process consists of generating the partial product's matrix, reducing the matrix to 2 rows followed by the final carry regarding the performance characteristics of a multiplier. Previously, partial product reduction has been achieved by using carry-save adders (CSA) consisting of rows of 3:2 compressors. Recently, a focus has been put on higher order reduction scheme mainly by using 4:2 compressors and it is constructed with 23-bit adder cells [3]. Several optimization approaches have been investigated and employed in [4–9] for 4-2 exact and approximate compressors [10]. In this work, we considered both the type of compressors and parameters like area and speed considered for VLSI implementation. Numerous multiplier architectures for approximate multiplication had been implemented in [11–14] and most of them used the truncation method. In [15], a new inaccurate multiplier architecture has been designed and validated for Arithmetic Data Value Speculation (ADVS). However, these works have involved in designing novel structures, investigating and implementing optimization methods for the reduction of partial product in multiplication without considering compressors. The metrics like PSNR and Normalized Mean Error Distance (NMED) are considered to evaluate the results of image multipliers. As like trade-off between various metrics in VLSI circuits, here the trade-off between precision and power is also quantitatively evaluated to find the suitability for multimedia applications.

This paper utilized two unique approximate 4-2 compressors and exact 5-2 compressors for partial product reduction in the multiplier. Then, these inaccurate data compressors are used in appropriate stages of Dadda multiplier and 4 novel architectures are designed for inexact multiplication. In new multiplier architectures, Sklansky Tree Adder (STA) is implemented at the last phase of addition to reduce the area and the overall latency.

2 Data Compressors

The fundamental concept of an n to 2 data compressor is, any number of input bits can be added to produce 2-bit outputs while separating sums and carries. This shows that all the input columns in the table can be added in parallel by omitting the result of an earlier column, generating 2-output adder with a small amount of delay which is independent of the input's size.

2.1 Exact 4 to 2 Compressor and 5 to 2 Compressor

Exact 4 to 2 Compressor (E42C): The conventional arithmetic circuits like CSA and parallel multiplication were designed to scale down the number of bits from n to 2; therefore, n-2 compressors have been generally implemented in computer arithmetic applications in computer arithmetic applications. Figure 1(a) demonstrates the basic version of E42C. An E42C is operated with 4 primary inputs from x_1 to x_4 plus one carry input C_{in} and produces 3 outputs such as C, $C_{\text{out,}}$ and S. So simply, it is a 5-bit adder. Here, two carry outputs C and C_{out} are carrying equal weightages. The C_{out} depends on only primary inputs and it forms the C_{in} of the next column. Therefore, it clearly shows that C_{out} is independent of C_{in} [10].

$$\text{Sum} = x_1 \oplus x_2 \oplus x_3 \oplus x_4 \oplus C_{\text{in}} \tag{1}$$

$$C_{\text{out}} = (x_1 \oplus x_2)x_3 + (x_1 \oplus x_2)' C_{\text{in}} \tag{2}$$

$$C = (x_1 \oplus x_2 \oplus x_3 \oplus x_4)C_{\text{in}} + (x_1 \oplus x_2 \oplus x_3 \oplus x_4)' x_4 \tag{3}$$

5 to 2 Compressor (52C): A widely used compressor of significant importance is the 52C for high precision and high-speed multipliers and it is illustrated in Fig. 1(b). It is working with 5 primary inputs (x_1–x_5) and 2 carry-in (C_{i1} and C_{i2}) from the earlier step. For the 7 input bits, the 52C produces sum, carry, and 2-carry out bits (C_{o1} and C_{o2}). The 52C generates an output of the same weight as the inputs, and three outputs, Carry, $C_{o1,}$ and C_{o2} and weighted one binary bit order higher. The outputs are given to the neighboring compressor of higher significance. All the 52Cs of different designs abide by Eq. (4). A new set of 52C implementations with an adjournment of 4 XORs is presented in [16]. A modified set of output expressions satisfying Eq. (4) are used in their design and they are shown in equations from (5) to (8).

$$x_1 + x_2 + x_3 + x_4 + x_5 + \text{C}_{i1} + \text{C}_{i2} = \text{sum} + 2(\text{carry} + \text{C}_{o1} + \text{C}_{o2}) \tag{4}$$

$$\text{Sum} = x_1 \oplus x_2 \oplus x_3 \oplus x_4 \oplus x_5 \oplus \text{C}_{i1} \oplus \text{C}_{i2} \tag{5}$$

$$\text{C}_{o1} = x_1 x_2 + (x_1 + x_2)x_3 \tag{6}$$

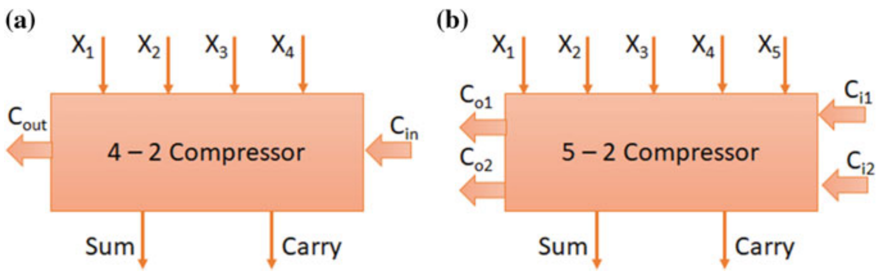


Fig. 1 Block diagram of **a** E42C, **b** 52C

$$C_{o2} = (x_4 \oplus x_5)C_{i1} + (x_4 \oplus x_5)x_3 \tag{7}$$

$$\begin{aligned}\text{Carry} = {} & ((x_1 \oplus x_2 \oplus x_3)(x_4 \oplus x_5 \oplus C_{i1}))C_{i2} \\ & + (\text{not}(x_1 \oplus x_2 \oplus x_3)(x_4 \oplus x_5 \oplus C_{i1}))(x_1 \oplus x_2 \oplus x_3)\end{aligned} \tag{8}$$

2.2 Approximate Compressors

In the truth table of an E42C, the carry output of 24 states out of 32 are having the same value of C_{in} and it helps for approximate design. In Approximate 4 to 2 Compressor I (A42CI), the carry is adjusted to C_{in} by changing the remaining 8 carry outputs. Here, carry output C is assumed as the higher weight of a binary bit and an incorrect value will generate a difference value of 2 in the output. For the input value 01001, the correct output is 010 that is equal to 2. By assuming the carry output C to C_{in}, the A42CI will generate the 000 at the output which is equal to 0. This difference value is inadequate for accurate computations but it can be paid by adjusting the C_{out} and sum values. In this case, we will consider the second half of Table 1. The sum value is simplified to 0 and it helps to minimize the difference between inexact and exact outputs. This kind of approach simplifies the design complexity with increased error rate and it is realized by comparing Eqs. (1–3) and (9–11).

$$C = C_{\text{in}} \tag{9}$$

$$\text{Sum} = C'_{\text{in}}(x_1 \oplus x_2)' + (x_3 \oplus x_4)' \tag{10}$$

$$C_{\text{out}} = x_1 \,.\, x_2 + x_3 \,.\, x_4 \tag{11}$$

Further to improve the performance of A42CI, the carry C and C_{out} equations are interchanged even they are having the same weight. This design is called as Approximate 4 to 2 Compressor II (A42CII). In this design, carry C uses the right-hand side of Eq. 6 and C_{out} is C_{in}. Initially, C_{in} value is 0 and C_{out} value is either 0 or 1. Then C_{out} and C_{in} value will be 0 in all stages so that they are discounted in our circuit realization. The A42CII has realized by the expressions in Eqs. 12 and 13.

$$\text{Sum} = (x_1 \oplus x_2 + x_3 \oplus x_4)' \tag{12}$$

$$C_{\text{out}} = (x_1 \,.\, x_2)' + (x_3 \,.\, x_4)' \tag{13}$$

For example, when all inputs are 1, the addition value is 4_{10}. However, the A42CII produces a 1 for the carry and sum. The base 10 representation of the outputs, in this case, is 3. This design has therefore 4 incorrect outputs out of 16 outputs and the difference value is presented in Table 2.

Table 1 Truth table for E42C and A42CI

C_{in}	Primary inputs				E42C			A42CI			
C_{in}	X_4	X_3	X_2	X_1	C_{out}	C	Sum	C_{out}	C	Sum	Diff.
0	0	0	0	0	0	0	0	0	0	1	1
0	0	0	0	1	0	0	1	0	0	1	0
0	0	0	1	0	0	0	1	0	0	1	0
0	0	0	1	1	1	0	0	0	0	1	−1
0	0	1	0	0	0	0	1	0	0	1	0
0	0	1	0	1	1	0	0	1	0	0	0
0	0	1	1	0	1	0	0	1	0	0	0
0	0	1	1	1	1	0	1	1	0	1	0
0	1	0	0	0	0	0	1	0	0	1	0
0	1	0	0	1	0	1	0	1	0	0	0
0	1	0	1	0	0	1	0	1	0	0	0
0	1	0	1	1	1	0	1	1	0	1	0
0	1	1	0	0	0	1	0	0	0	1	−1
0	1	1	0	1	1	0	1	1	0	1	0
0	1	1	1	0	1	0	1	1	0	1	0
0	1	1	1	1	1	1	0	1	0	1	−1
1	0	0	0	0	0	0	1	0	1	0	1
1	0	0	0	1	0	1	0	0	1	0	0
1	0	0	1	0	0	1	0	0	1	0	0
1	0	0	1	1	1	0	1	0	1	0	−1
1	0	1	0	0	0	1	0	0	1	0	0
1	0	1	0	1	1	0	1	1	1	0	1
1	0	1	1	0	1	0	1	1	1	0	1
1	0	1	1	1	1	1	0	1	1	0	0
1	1	0	0	0	0	1	0	0	1	0	0
1	1	0	0	1	0	1	1	1	1	0	1
1	1	0	1	0	0	1	1	1	1	0	1
1	1	0	1	1	1	1	0	1	1	0	0
1	1	1	0	0	0	1	1	0	1	0	−1
1	1	1	0	1	1	1	0	1	1	0	0
1	1	1	1	0	1	1	0	1	1	0	0
1	1	1	1	1	1	1	1	1	1	0	−1

Table 2 Truth table for A42CII

X_4	X_3	X_2	X_1	C	Sum	Difference
0	0	0	0	0	1	1
0	0	0	1	0	1	0
0	0	1	0	0	1	0
0	0	1	1	0	1	−1
0	1	0	0	0	1	0
0	1	0	1	1	0	0
0	1	1	0	1	0	0
0	1	1	1	1	1	0
1	0	0	0	0	1	0
1	0	0	1	1	0	0
1	0	1	0	1	0	0
1	0	1	1	1	1	0
1	1	0	0	0	1	−1
1	1	0	1	1	1	0
1	1	1	0	1	1	0
1	1	1	1	1	1	−1

3 Tree Structure Adders

The low-cost multipliers can be constructed by reducing the partial product sizes. The tree adders address the solution for power consumption and latency in arithmetic circuits. The tree adders are developed from the computation concept of Carry-Lookahead (CLA) adders and it is viewed as parallel CLA. These structures target high-performance applications. In this work, the Sklansky Tree Adder (STA) is considered because it reduces power consumption of adder circuits and increases speed than other tree adders in the earlier works. The binary trees of cells in the STA first generate all possible carry inputs at the same time and it follows a systematic configuration. The STA ultimately reduces the delay to logarithmic of number of bits (number of bits $= N$) by calculating the midway prefixes with the large group prefixes. It increases the number of fan-outs at each stage. In this adder, a binary tree of propagate and generate cells will first simultaneously generate all the carries and C_{in}. It constructs 2-, 4-, 8-bit adders and so on by adjoining 2 smaller adders (2-bit adders), respectively [17].

4 Hybrid Multipliers

This work considered Dadda multiplier because it requires a minimum number of adder stages for the summation of partial products. The design of Dadda multiplier

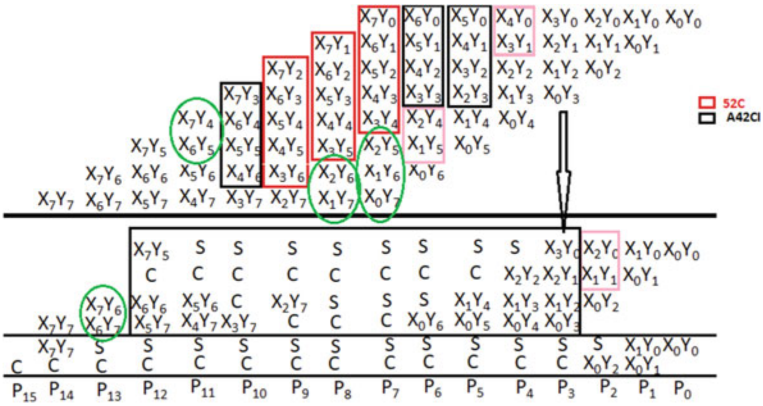


Fig. 2 Reduction structure of partial product in HyMUL 1

comprises three steps and they are followed. The partial product is generated in the first step and it is reduced to a height of 2 in the second step. At last step, these 2-rows are combined by using various addition algorithms. Initially, 8×8 unsigned Dadda multiplier is considered to evaluate the performance of compressors in the multiplier. In this work, all the partial products are generated using AND gates and inexact compressors discussed in this paper are employed at the addition portion to minimize the partial products. The last part is designed with STA structure to calculate the final product value. The various stages of HyMUL1 for $n = 8$ is shown in Fig. 2. The reduction part uses adders and compressors. In this paper, five cases are considered for designing the inexact multipliers and they are named as HyMUL 1, HyMUL 2, HyMUL 3, HyMUL 4 and HyMUL 5. The feature of each multiplier is given in Table 2. (These multipliers are called as approximate multiplier because the multiplier is designed with approximate compressors.) For understanding, HyMUL 1 is explained here with its feature. The hybrid multiplier 1 (HyMUL 1) is designed with the approximate compressor I and exact 5-2 compressor to minimize the partial product hight.

4.1 HyMUL 1

A42CI is used for all 4-variable additions and 52C is used for all 5-variable additions. The height of subsequent stages $= 2(X - i)$, where $2X$ is a nearest smaller integer to N; i ranges from 0 to $X-1$. The stage heights are stage $2 = 2X$, Stage $3 = 2X - 1$, Stage $4 = 2X - 2\ldots$ until final stage height is 2. This setup is achieved with the usage of 52C while maintaining stage height. The 8×8 hybrid multiplier 1 architecture is presented in Fig. 2. The maximum height of the partial products stage (Stage1) is 8 bits. The nearest 2 m integer smaller than 8(23) is 4, i.e., 2(2-0). So, the height

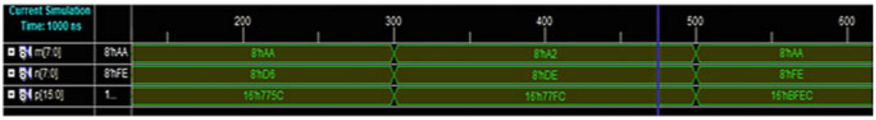


Fig. 3 Simulation result of HyMUL 2

Table 3 Area and delay comparison of various multipliers

Multipliers	Features	CPA		STA	
		Area	Delay (ps)	Area	Delay (ps)
Multiplier 1	A42CI in all columns	858	22.077	801	20.489
Multiplier 2	A42CII in all Columns	738	22.42	716	19.921
Multiplier 3	E42C in all columns	807	22.654	784	21.638
HyMUL 1	A42CI + 52C	840	21.739	817	19.95
HyMUL 2	A42CII + 52C	718	21.448	693	18.721
HyMUL 3	A42CI in LSB + 52C in MSB	810	21.934	802	20.351
HyMUL 4	A42CII in LSB + 52C in MSB	775	20.756	768	20.478
HyMUL 5	E42C + 52C	786	20.324	778	20.324

of Stage 2 should be 4 bits, which is maximum column (C) height. This hybrid multiplier 1 uses 14 approximate compressor I structures, 3 exact 5-2 compressor, 3 half adders, and 18 full adders. The usage of STA at the last portion of multipliers, the required number of 3-bit adders can be minimized.

5 Simulation Results

The multipliers have designed using VHDL and simulated with Xilinx ISE Simulator. Figure 3 demonstrates the simulation result of HyMUL 2 for 8 × 8. The area in terms of number transistors and delay (in ps) are tabulated and it is presented in Table 3. The HyMUL 2 has less transistor count compared to other multipliers. The transistor count is used in this paper as metric of area occupancy. And HyMUL 3 also consumes less hardware area with lower delay parameter. From Table 3, it is clearly understood that STA at the last portion of the proposed hybrid multipliers helps to reduce the latency during the computation process. Among all multipliers, the HyMUL 2 which is designed using A42CII and 52C in all columns of partial products has less delay as shown in Table 3.

Fig. 4 Example 1: multipliers for image processing

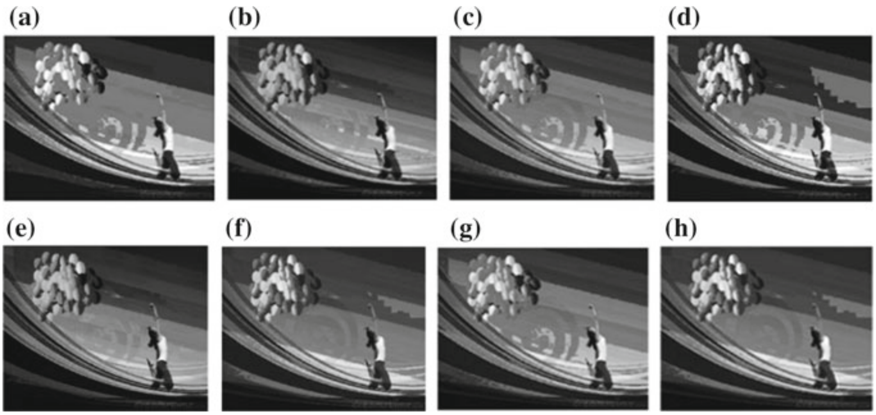


Fig. 5 Image multiplication results for Example 1

6 Hybrid Multipliers for Image Processing

The multipliers are applied to image processing applications like image manipulations or image overlapping. A multiplier for image processing can be designed to multiply the pixels of two images. This multiplication process is actually a pixel-by-pixel multiplication, thus two images are overlapped. Figure 4 shows an example of two input images and the resulting output image is provided. The image multiplication results of examples are given in Figs. 5 and 6.

6.1 Performance Analysis

PSNR: PSNR that is based on the Mean Squared Error (MSE) are computed to evaluate the quality of the output image. The equations for the MSE and PSNR are presented and discussed in [1]. Table 4 shows the comparison of PSNR of the output images generated by all the multipliers. The HyMULs are designed using various combinations of A42CI, A42CII, E42C, and 52C compressors and produces incorrect value. However, when applied to image processing applications, these multipliers

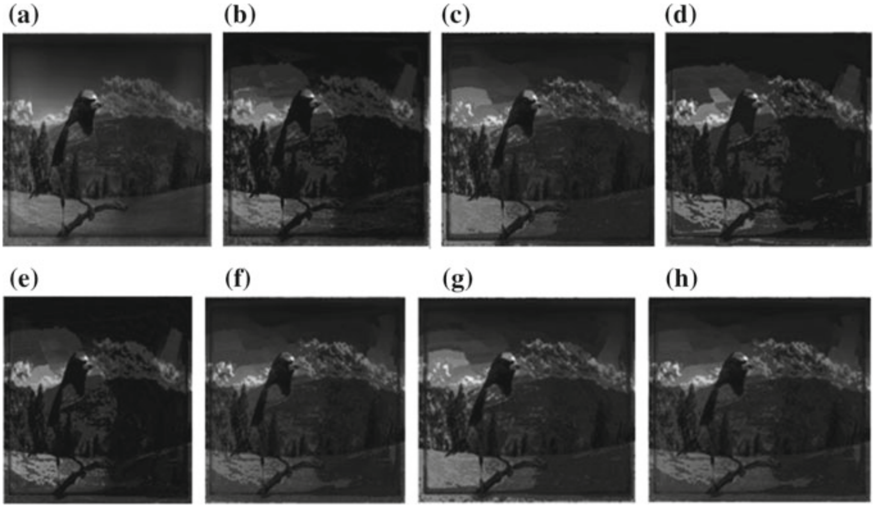


Fig. 6 Image multiplication results for Example 2 (for Figs. 5 and 6: **a** Multiplier 1, **b** Multiplier 2, **c** Multiplier 3, **d** HyMUL 1, **e** HyMUL 2, **f** HyMUL 3, **g** HyMUL 4, **h** HyMUL 5)

Table 4 PSNR and NMED comparison of all the multipliers with STA

Multipliers	PSNR	NMED
Multiplier 1	26.6688	3.0942
Multiplier 2	22.3255	7.4891
Multiplier 3	29.111	5.0204
HyMUL 1	45.2801	0.0757
HyMUL 2	49.7683	0.0709
HyMUL 3	38.8971	0.4601
HyMUL 4	41.7465	0.4263
HyMUL 5	37.9894	0.5728

results in improved peak signal-to-noise ratio (PSNR) value as shown in Table 4. The proposed multiplier HyMUL 2 of examples is found to have high PSNR value when compared to other multipliers and its value is around 50 dB which is applicable for most applications.

NMED: The Error Distance (ED) and NMED are considered to evaluate the performance of approximate arithmetic circuits for image processing applications. For multiplication, ED is the arithmetic difference between the exact product (M) and the approximate product (M'). The MED is computed by taking the average of EDs for a set of outputs (for the set of inputs). The error distance is measured between for various multipliers are listed in Table 4. The equations for ED, MED and NMED are presented in [1] and simulated using MATLAB.

7 Conclusions

The multipliers are considered as an important module of processors and DSPs so that various algorithms like convolution and filtering for image and signal processing can be implemented. The computer arithmetic supports various schemes and structures for both exact and inexact computation. In this work, five different schemes are presented in 8-bit Dadda multiplier using A42CI, A42CII, and 52C for only inexact computation. On the experimental results, HyMUL 2 occupied a smaller hardware area and its latency is considerably reduced than all other multipliers. The multipliers designed using the various combinations of A42CI, A42CII, E42C and 52C results in the inexact final product but when applied to image processing applications, image multipliers result in high PSNR and low NMED values compared to conventional or exact multipliers. And in future, higher order multiplier design with approximately 5-2 compressor can be designed to reduce the delay and transistor count further.

References

1. Momeni JH, Montuschi P, Lombardi F (2014) Design and analysis of approximate compressors for multiplication. IEEE Trans Comput 64(4):984–994
2. Han J, Orshansky M (2013) Approximate computing: an emerging paradigm for energy-efficient design. In: IEEE ETS, pp 1–6
3. Chang P, Ahmadi M (2009) High speed low power 4:2 compressor cell design. In: International symposium on signals, circuits and systems, Iasi, pp 1–4
4. Chang C, Gu J, Zhang M (2004) Ultra Low-voltage low- power CMOS 4-2 and 5-2 Compressors for fast arithmetic circuits. IEEE Trans Circuits Syst 51(10):1985–1997
5. Gu J, Chang CH (2003) Ultra low-voltage, low-power 4-2 compressor for high speed multiplications. In: 36th IEEE international symposium circuits systems, Bangkok, Thailand
6. Margala M, Durdle NG (1999) Low-power low-voltage 4-2 compressors for VLSI applications. In: IEEE Alessandro volta memorial workshop low-power design, pp 84–90
7. Parhami B (2010) Computer arithmetic: algorithms and hardware designs, 2nd edn. Oxford University Press, New York
8. Prasad K, Parhi KK (2001) Low-power 4-2 and 5-2 compressors. In: 35th Asilomar conference on signals, systems and computers, vol 1, pp 129–133
9. Ercegovac MD, Tomas L (2003) Digital arithmetic. Elsevier, Amsterdam
10. Ma J, Man K, Krilavicius T, Guan S, Jeong T (2011) Implementation of high performance multipliers based on approximate compressor design. In: International conference on electrical and control technologies (ECT), pp 96–100
11. Mahdiani HR, Ahmadi A, Fakhraie SM, Lucas C (2010) Bio-inspired imprecise computational blocks for efficient VLSI implementation of soft-computing applications. IEEE Trans Circ Syst I: Regul Pap 57(4):850–862
12. Schulte MJ, Swartzlander EE (1993) Truncated multiplication with correction constant. In: IEEE workshop on VLSI signal processing VI, pp 388–396
13. King EJ, Swartzlander EE (1998) Data dependent truncated scheme for parallel multiplication. In: 31st Asilomar conference on signals, circuits and systems, pp 1178–1182
14. Kulkarni P, Gupta P, Ercegovac MD (2011) Trading accuracy for power in a multiplier architecture. J Low Power Electron 7(4):490–501
15. Kelly D, Phillips B, Al-Sarawi S (2009) Approximate signed binary integer multipliers for arithmetic data value speculation. In: Conference on design and architectures for signal and image processing, pp 97–104

16. Menon R, Rdhakrishnan D (2006) High performance 5:2 compressor architectures. In: IEE Proceeding of circuits devices systems, vol 153, no 5
17. Sankar DR, Ali SA (2013) Design of Wallace tree multiplier by Sklansky adder. Int J Eng Res Appl 3(1):1036–1040

Intervertebral Disc Classification Using Deep Learning Technique

J. V. Shinde, Y. V. Joshi and R. R. Manthalkar

Abstract This paper describes the semiautomatic method for diagnosis of intervertebral disc degeneration according to Pfirrmann's five scale (1–5) grading system, which is used in the assessment of disc degeneration severity. Total 1123 discs are obtained after augmentation from 120 subject's T2-weighted lumbar scans. Manual classification into five grades is done by experts. Our method is extracting 59 features using Local Binary Pattern for texture analysis and 4096 features using pretrained CNN. 1×59 and 1×4096 feature vectors are fused to form 1×4155 feature vector to train our multiclass Support Vector Machine classifier. This feature level fusion method is able to achieve 80.40% accuracy. A Quantitative analysis is done using parameters, viz.,—Accuracy, Sensitivity, Specificity, Precision, Recall, F1 score, etc.

1 Introduction

Spinal degeneration [1–3] is commonly found in two anatomical parts called vertebra and intervertebral disc that results in mild to severe back pain. Degenerative lumbar spine diseases include disc herniation and Modic changes which are highly associated with lower back pain. Modic changes and different types of disc herniation [4] are clearly visible on Magnetic Resonance Imaging (MRI). In clinical evaluation disc degeneration, severity is assessed with mid-sagittal slice of T2-weighted images. It provides both internal chemical composition and structural integrity information

J. V. Shinde (✉)
Department of Computer Engineering, L.G.N. Sapkal College of Engineering, Nasik, India
e-mail: jv.shinde@rediffmail.com

Y. V. Joshi · R. R. Manthalkar
Department of Electronics & Telecommunication, S.G.G.S Institute of Engineering and Technology, Nanded, India
e-mail: yvjoshi@sggs.ac.in

R. R. Manthalkar
e-mail: rrmanthalkar@sggs.ac.in

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_55

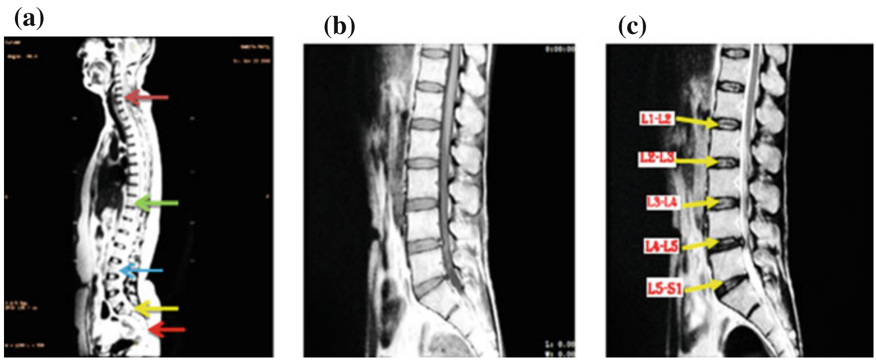


Fig. 1 **a** Whole spine image, Brown arrow cervical, Green arrow Thoracic, Blue arrow Lumbar, Yellow arrow Sacrum and Red arrow shows coccyx section, **b** mid-sagittal T1—weighted image and **c** T2-weighted image with 5 disc positions in between vertebras. MR Image source: Samarth Diagnostic center, Nasik (M.S), India

about disc tissue. Biochemical changes are reflected in MRI signal intensity and compared with normal levels.

In addition, T2 image provides useful information like disc shape, size, water content, erosion of vertebral endplates, wear and tear of the outer wall of disc, etc., which allows investigating various disc disorders like disc narrowing and disc herniation. Pfirrmann's five scale (1–5) grading system is very popular in Magnetic Resonance Imaging assessment [5].

Spine is made up of small bones called vertebra, muscles, ligaments, nerves, and intervertebral discs. The whole spinal column is divided into 5 sections namely cervical, thoracic, lumbar, sacrum, and tiny coccyx. The cervical section contains 7 vertebrae C1–C7 and discs, Thoracic contains 12 vertebrae T1–T12 and discs, lumbar section contains 5 vertebrae L1–L5 and 5 discs, sacrum shows fused vertebrae S1–S5 which articulates with the hip bone of pelvis, spinal column is terminated by tiny coccyx consist of fused vertebrae shown in Fig. 1.

Commonly used pulse sequences are T1, T2-weighted and T2-weighted with fat saturation (STIR) technique

T2-weighted mid-sagittal sequence images are used for assessing Intervertebral Discs (IVD), which provide better visualization to morphological- and tissue-related biochemical changes. These changes are reflected in the image in terms of variations in MR signal intensities and irregularities in the shape and size of the disc. Alterations in intervertebral disc height, signal intensity, and the distinction between center part nucleus pulposus and outer wall annulus fibrosus are major signs of disc degeneration. All these features are evaluated for determining disc degeneration severity. In the literature, various grading schemes have been proposed to assist clinicians [6, 7]. Figure 2 shows Pfirrmann's grading scheme (1–5).

In Grade-1 class, the disc looks normal and a clear difference can be observed in NP and AF. Also, the height of the disc seems to be normal. NP reflects in a

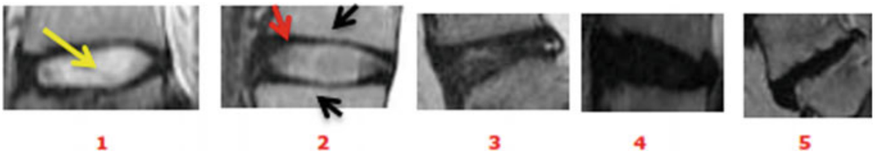


Fig. 2 1. Normal Disc 2. Normal Disc 3. Mild Degeneration 4. Moderate Degeneration 5. Severe Degeneration. Yellow arrow shows NP, Black arrows shows vertebra and Red arrow shows AF

brighter white shade. Grade-2 class indicates a normal disc with normal height but NP looks inhomogeneous with or without horizontal bands. Grade-3 class shows mild degeneration and NP is Inhomogeneous, Gray with slightly decreased disc height. There is an unclear distinction between NP and AF. If disc height is moderately decreased and the distinction between NP and AF is lost, NP looks inhomogeneous, Gray to black then disc falls into Grade-4 class. Grade-5 belongs to the class of severe degeneration, where disc height is totally collapsed and NP is black. Also, the distinction between NP and AF is completely lost.

2 Related Work

Intervertebral Disc Degeneration (IVDD) is becoming a universal health issue, which results in tremendous economic losses and damage to productivity every year. IVD degeneration is a common cause of lower back pain found commonly in almost every person in late adulthood to a varying degree.

For surgical removal of disc radiological scoring systems and histological systems are utilized to assess the progression of intervertebral disc degeneration. MRI helps to record signal variations in the disc while microstructural changes are less appreciated. MRI also helps to find the changes in the endplate of vertebra known as the Modic changes.

2.1 Lumbar Disc Degeneration Scoring Systems Utilized in Following Imaging Modalities

2.1.1 X-ray Imaging

There are totally six classification systems are developed and used for X-ray diagnosis, Kellgren's [8] 0–2 scale is based on movement, sclerosis, lipping, and spacing criteria. Gordon [9] proposed 1–4 scale scoring based on sclerosis, narrowing, osteophytes. Lane [10] has been developed 0–3 scale scoring method using criteria joint space narrowing, osteophytes and sclerosis. Mimura [11] developed 1–4 scale using disc height, osteophytes, endplate, and sclerosis. Taking into account the disc height, osteophytes, endplate sclerosis, Schmorl's nodes, and vacuum phenomenon Madan [12] developed a 0–3 scale classification system.

Thalgott [13] used A–F scale scoring using both X-ray and MRI on sclerosis and osteophytes criteria. Wilke [14] in (2006) proposed 0–3 scoring on height loss, osteophyte formation, and diffuse sclerosis.

2.1.2 Magnetic Resonance Imaging

Eight classification systems have been proposed for IVD degeneration using MRI. One of the strong radiological grading systems proposed by Pfirrmann et al. [5] is based on morphological alterations found in the disc. Pfirrmann's 1–5 grades are shown in Fig. 2. Schneiderman [15] formed 1–4 scales based on disc height, signal intensity, and patterns in MRI. Butler [16] proposed 1–4 scaling system based on criteria borders of annulus and nucleus, disc space, herniation, signal intensity. Tertti's [17] 1–3 scale is based on only signal intensity variations found in the sagittal plane.

Gunzburg [18] proposed 0–3 scale scoring observing only central part that is nucleus signal. Disc space narrowing and signal intensity are observed in all three planes namely Axial, Sagittal, and Coronal by Southern et al. [19] to propose a 1–4 scaling system. Askar [20] has proposed 1–4 grading using both axial and sagittal planes based on criteria like disc height reduction, radial tears, and annular's shape. Griffith's [21] 1–8 scoring system is based on rim lesions, annular tears, bulging, disc height and difference between NP and AF.

2.1.3 Histological Grading

Histological grading involves microscopic tissue level investigation. As per Boos et al. [22] criteria four parameters are evaluated at the microscopic level. Chondrocyte proliferation, Tears of clefts, granular changes and mucoid changes with 0–22 scale. Three more grading systems are presented: Gunzburg [18] proposed 0–3 scale scoring, Berlemann [23] proposed 1–4 scale scoring, and Weiler [24] proposed 0–15 scale scoring.

Many contributions have been made in literature for automatic diagnosis/classification of a number of spinal degeneration conditions like Modic changes and intervertebral disc degeneration. Methods proposed by Corso et al. [25] and Alomari et al. [26–28] deals with binary classification on the basis of absence or presence of abnormality like disc desiccation/degeneration taken in consideration the appearance and shape of disc. The various intensity and texture features are extracted from ROI and shape information are evaluated using ASM and GVF-snake-based segmentation algorithms. One of their contributions is using a probabilistic model for automatically labeling the disc and extracting ROI using ASM. Classification accuracy is compared using heterogeneous classifiers. But radiological grading is not performed.

Instead of using conventional T2 weighted MR images for classification of intervertebral discs, Watanabe et al. [29] proposed a system which is very useful in the detection of early degeneration in the disc using axial T2 mapping technique. Using human cadaveric disc specimens Southern et al. found a correlation between MRI and quantitative discomanometry in the assessment of lumbar disc degeneration [19].

Marcelo et al. [30] developed a semiautomatic method for diagnosis of intervertebral disc degeneration using T2-weighted lumbar images according to Pfirmann's five scale grading system. Each ROI is detected using a binary mask followed by Haralick feature extraction for texture analysis using gray level co-occurrence matrix. Disc classification is done using a multilayer perceptron artificial neural network with full attribute vectors.

Using Pfirmann's scoring system, automatic classification of intervertebral disc degeneration using T2-weighted images is performed by Castro-Mateos et al. [31]. The region of interest is segmented using an active contour model. Using shape and intensity features of intervertebral disc Neural Network is trained to achieve classification task.

Most of the spinal disease diagnosis studies have been primarily conducted with handcrafted features. A very few approaches utilize deep learning based frameworks for classification of IVD degeneration. The Pretrained Convolution Neural Networks (CNN) has been utilized mainly in medical diagnosis for pathology detection.

One successful example based on CNN is 'U-net' a medical image segmentation framework developed by Ronneberger et al. [32]. It works on few training image dataset and yields more precise segmentation.

Pfirmann's grading/scoring system is used by the method proposed by Jamaludin et al. [33] their scheme automatically predicts radiological scoring as well as disc narrowing and marrow changes from multiple slice lumbar spine MR images. Their CNN based framework is modified version of VGG-M known by name 'SpineNet'. İt is employed to detect, localize, and classify multiple abnormalities of spinal objects like vertebra and discs in T2-weighted MRI.

3 Proposed Methodology

We have proposed feature-level fusion approach for IVD classification based on Pfirrmann's grading scale. The procedure of the proposed algorithm is described as follows (Fig. 3).

3.1 Image Preprocessing

T2-weighted images are manually cropped and five discs from lumbar sections are saved separately. Categorizations of those discs are done in five classes as per the

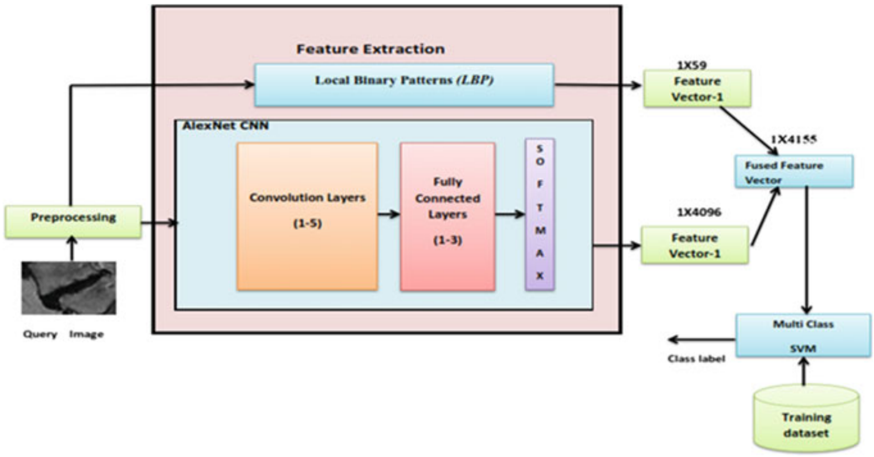


Fig. 3 Architecture diagram of classification process using feature level fusion

recommendation of experts. Images are preprocessed by using super-resolution [34] technique to enhance the low-resolution images.

3.2 *Local Binary Patterns*

Local binary pattern [35] is computationally simple and an efficient texture operator used on spatial domain images in texture analysis applications. It uses grayscale image for texture classification as it is very robust to monotonic grayscale variations, which exist due to illumination variations. When LBP operator is applied on an image a binary number is obtained as a result. LBP representations can be expressed by following Eqs. (1) and (2). Image pixel labels are provided after thresholding the 3 × 3 neighborhood with a center pixel value.

The value of the LBP code of pixel (x_c, y_c) is given by

$$\mathrm{LBP}_{N,R} = \sum_{n=0}^{N-1} s_x(g_n - g_c)2^n \tag{1}$$

where N = no. of sampling points, R = radial distance, and g_c = center pixel

$$s_x(\mathrm{diff}) = \begin{cases} 1, & (\mathrm{diff}) \geq 1 \\ 0, & (\mathrm{diff}) < 0 \end{cases} \tag{2}$$

Following Fig. 4 shows the LBP code computation process.

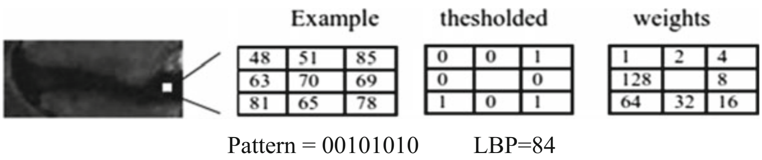


Fig. 4 Local binary pattern process

3.3 Convolution Neural Networks

Convolution Neural Networks (CNN) consist of two types of layers: convolution and pooling layers. In convolution layer entire image is convolved with the set of learned filters, f_n to get n corresponding feature maps which are then subsampled by average or max in pooling layer. It makes resulting feature invariant to small deformation and translation.

CNNs are used for improving the accuracy of image classification. It processes multiple array data. In last few years pretrained network models have gained popularity and widely used in medical imaging domain. 'AlexNet' [36] is one such pretrained model available for feature extraction. AlexNet contains 25×1 layer array with a series of five convolution layers, three fully connected layers followed by single softmax layer. A series of convolution layers are intermixed with rectified linear units (ReLU) and max-pooling layers. Classification is done by the final layer and its properties are depending on the classification task. This net can process RGB images that are $227 \times 227 \times 3$ in size.

3.4 Data Augmentation

In order to maximize the classification accuracy and to increase the number of samples available for the training. We have applied rotations by −2 and +2 degrees. Augmentation samples are used as extra samples in training. Finally, the number of class-wise samples is as follows.

Class-1 → 210 samples, Class-2 → 223 samples, Class-3 → 212 samples, Class-4 → 261 and Class-5 → 217 samples. In total, we have 1123 samples. From these 1123 samples, our system will select randomly at runtime 60% training and 40% testing samples. That results in overall 674 samples for training and 449 for testing.

3.5 Support Vector Machine Classifier

Support Vector Machine (SVM) is a supervised non-probabilistic binary linear classifier [37] widely used in classification and pattern recognition problems. To per-

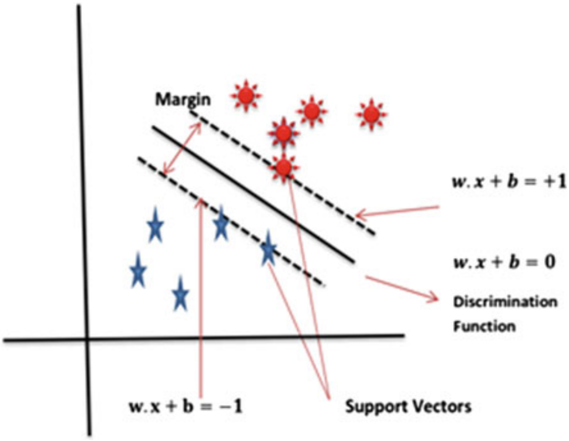


Fig. 5 Support Vector Machine for linear separable data

form classification SVM constructs decision boundaries called as hyperplanes in a multidimensional space that separates a set of objects belonging to different class memberships. The optimal hyperplane can be found by using feature points called support vectors (Fig. 5).

SVM training can be considered as the constrained optimization problem which minimizes the structural risk while maximizing the width of the margin.

$$\min(w, b) \quad \frac{1}{2}a^T a + C\sum_{i=1}^{N} \varepsilon_i \tag{3}$$

Subject to

$$y_i\left(W^{\mathrm{T}}\varphi(x_i) + b\right) \geq 1 - \varepsilon_i, \quad \varepsilon_i \geq 0, \forall_{i,} \tag{4}$$

where b is the bias, C is the tradeoff parameter, ε_i variable measures the deviation of a data point from the ideal condition and $\varphi(.)$ is the feature vector. A Support Vector Machine classifier algorithm constructs a model that allocates new sample to one of the two variable regions. Classification of the test sample p is performed by

$$y = sgn\left(\sum_{sv=1}^{N_{sv}} \propto_{sv} y_{sv} K(x_{sv}, p)\right) \tag{5}$$

Mapping of input samples into higher dimensional space is performed by the kernel function $K(x_{sv}, p)$.

3.6 Imaging Data Details

Mid-sagittal DICOM MR images are acquired using 1.5 T Siemens scanner from Samarth diagnostic center Nasik, India for 120 individuals of both sexes with age varying from 20 to 89 yrs. Imaging protocol as follows with parameters.

- T2-weighted: FRFSE SAG TR = 4220 ms, TE = 105 ms,
  Slice thickness = 4 mm, image matrix size is 512×512.
- T1-weighted: FRFSE SAG TR = 700 ms, TE = 114.34 ms,
  Slice thickness = 4 mm, image matrix size is 512×512.

T2-Weighted images are used for experimentation.

3.7 Implementation

Algorithm 1: Proposed Training Algorithm

Input: Train image, LBP block size.
Output: Fused feature vector for Train input image
Algorithm steps
Step 1: Image preprocessing (image resolution enhancement)
1.1 Read the train image **I** of size 227x227x3
1.2 Enhance resolution of **I** by applying Super resolution technique.
1.3 Refer enhanced image as **I_Sr**
Step 2:
Pass **I_Sr** image as an input to **CNN** to get feature vector **f1_cnn** of size **1x4096.**
Step 3: Computation of LBP feature vector
J = Resize **(I)** to **128x256**
3.1 Convert image **J** into blocks with block size **B_size.**
for every pixel in the block **do**
3.2 Perform computations to obtain LBP representation
3.3 Computation of histogram **hist_i** for each block **i**
3.4 All histograms are concatenated to obtain a feature vector **f2_lbp** of **1x59** size.
Step 4: Perform feature fusion to obtain fused feature vector.
Concatenate **f1_cnn** & **f2_lbp** to yield fused feature vector **f_fused** of **1x4155** size.
Step 5. Repeat Step 1 to 4 for all training images
Save feature vector in training database **F_train** matrix

Table 1 5 × 5 confusion matrix in the IVD classification

Correct class	Classified as				
	1	2	3	4	5
1	62	16	6	0	0
2	0	71	18	0	0
3	0	20	51	12	2
4	0	1	11	91	1
5	0	0	0	1	86

Algorithm 2: Proposed Testing Algorithm

Input: Test image sample
Output: class label for test image
Algorithm steps
Step 1: Perform step 1 through 4 of Algorithm 1 to obtain fused feature vector **F_test**.
Step 2: Multiclass Support Vector Machine classification
2.1. Compare the **F_test** feature vector against feature vector matrix **F_train**.
2.2. Find the probability of particular class using one against all model to find maximum probability.
Class Label having maximum probability will be the output label.

---

4 Results and Discussion

İn medical image classification many handcrafted feature descriptors are used, LBP is a texture descriptor having high discriminative power and invariant to grayscale and illumination changes, so widely used in medical image analysis.

While using deep learning methods very complex functions can be learned. İn this work we propose feature level fusion process where 59 features are extracted from LBP and total 4096 features are extracted from CNN. By fusion, we got 1 × 4155 size feature vector. SVM training is done and classification is performed by the same classifier using one against all model. Parameters Accuracy, Sensitivity, Specificity, Precision, Recall, and F1-Score are reported in Table 2. The resultant confusion matrix is 5 × 5 is shown in Table 1.

Table 2 Quantitative Analysis Accuracy, Sensitivity, Specificity, Precision, Recall, and F1-score

Class	Accuracy	Sensitivity	Specificity	Precision	Recall	F1-score
Class-1	0.95	0.74	1.00	1.00	0.73	0.84
Class-2	0.87	0.80	0.89	0.65	0.79	0.72
Class-3	0.84	0.60	0.90	0.59	0.60	0.59
Class-4	0.94	0.88	0.96	0.87	0.88	0.87
Class-5	0.99	0.98	0.99	0.96	0.98	0.97

5 Summary and Conclusion

İn this paper, we present a feature fusion method for intervertebral disc degeneration classification based on Pfirrmann's 1–5 scale. This method takes advantages of both feature extraction methods from the spatial domain using LBP and CNN form neural network domain. Accuracy is calculated as 80.40%. We are working on the relatively small dataset, but deep networks provide good results on a huge dataset. Our future study will be on the huge dataset and will try different CNN models for comparisons. In addition, dimensionality reduction techniques can be applied to achieve maximum accuracy.

Acknowledgements We thank Dr. Hemant Borse, Consultant Radiologist, Samarth Diagnostic Center, Nasik and Dr. Rajesh Jawale, Consultant Radiologist, Wockhardt Hospital Nasik (M.S.) India who provided insight and expertise that greatly assisted the research. We are immensely grateful to Dr. Hemant Borse and his technical team for providing spine MR image dataset for research work.

References

1. Bindra S, Sinha AGK, Benjamin AI (2015) Epidemiology of low back pain in indian population: a review. Int J Basic Appl Med Sci 5(1):166–179. ISSN 2277–2103 (Online) 015
2. Fardon DF (2014) Lumbar disc nomenclature: version 2.0 Recommendations of the combined task forces of the North American Spine Society, the American Society of Spine Radiology and the American Society of Neuroradiology Review Article. Spine J 14:2525–2545
3. Nazeer M, Rao SM, Soni S, Ravinder M (2015) Lower back pain in South Indians: causative factors and preventive measures. Sch J App Med Sci 3(1D):234–243. ISSN 2347-954X
4. Modic MT, Ross JS (2007) Lumbar degenerative disk disease. Radiology 245:43–61
5. Pfirrmann CWA, Metzdorf A, Zanetti M, Hodler J (2001) Magnetic resonance classification of lumbar intervertebral degeneration. SPINE 26:1873–1878
6. Ketler A, Wilke HJ (2006) Review of existing grading systems for cervical or lumbar disc and facet joint degeneration. Eur Spine J 15:705–718
7. Bennekar LM, Heini PF, Anderson SE (2005) Correlation of radiographic and MRI parameters to morphological and biochemical assessment of intervertebral disc degeneration. Eur Spine J 14:27–35
8. Kellgren JH, Lawrence JS (1952) Rheumatism in miners Part II X-ray study. Br J Ind Med 9(3):197–207

9. Gordon SJ, Yang KH, Mayer PJ, Mace AH Jr, Kish VL, Radin EL (1991) Mechanism of disc rupture. A preliminary report Spine (Phila Pa 1976) 16(4):450–456
10. Lane NE, Nevitt MC, Genant HK, Hochberg MC (1993) Reliability of new indices of radiographic osteoarthritis of the hand and hip and lumbar disc degeneration. J Rheumatol 20(11):1911–1918
11. Mimura M, Panjabi M, Oxland T, Crisco J, Yamamoto I, Vasavada A (1994) Disc degeneration affects the multidirectional flexibility of the lumbar spine. Spine (Phila Pa 1976) 19(12):1371–1380
12. Madan SS, Rai A, Harley JM (2003) Interobserver error in interpretation of the radiographs for degeneration of the lumbar spine. Iowa Orthop J 23:51–56
13. Thalgott J, Albert T, Vaccaro A, et al (2004) A new classification system for degenerative disc disease of the lumbar spine based on magnetic resonance imaging, provocative discography, plain radiographs and anatomic considerations. Spine J 4(6 Suppl):S167
14. Wilke HJ, Rohlmann F, Neidlinger-Wilke C, Werner K, Claes L, Kettler A (2006) Validity and interobserver agreement of a new radiographic grading system for intervertebral disc degeneration: Part I. Lumbar spine. Eur Spine J 15(6):720–730
15. Schneiderman G, Flannigan B, Kingston S, Thomas J, Dillin WH, Watkins RG (1987) Magnetic resonance imaging in the diagnosis of disc degeneration: correlation with discography. Spine (Phila Pa 1976) 12(3):276–281
16. Butler D, Trafimow JH, Andersson GB, McNeill TW, Huckman MS (1990) Discs degenerate before facets. Spine (Phila Pa 1976) 15(2):111–113
17. Tertti M, Paajanen H, Laato M, Aho H, Komu M, Kormano M (1991) Disc degeneration in magnetic resonance imaging. A comparative biochemical, histologic, and radiologic study in cadaver spines. Spine (Phila Pa 1976) 16(6):629–634
18. Gunzburg R, Parkinson R, Moore R, Cantraine F, Hutton W, Vernon Roberts B, Fraser R (1992) A cadaveric study comparing discography, magnetic resonance imaging, histology, and mechanical behavior of the human lumbar disc. Spine 17(4):417–426
19. Southern EP, Fye MA, Panjabi MM, Patel TC, Cholewicki J (2000) Disc degeneration: a human cadaveric study correlating magnetic resonance imaging and quantitative discomanometry. Spine 25(17):2171–2175
20. Askar Z, Wardlaw D, Muthukumar T, Smith F, Kader D, Gibson S (2004) Correlation between intervertebral disc morphology and the results in patients undergoing Graf ligament stabilisation. Eur Spine J 13(8):714–718
21. Griffith JF, Wang YX, Antonio GE, Choi KC, Yu A, Ahuja AT, Leung PC (2007) Modified Pfirrmann grading system for lumbar intervertebral disc degeneration. Spine (Phila Pa 1976) 32(24):E708-E712
22. Boos N, Weissbach S, Rohrbach H, Weiler C, Spratt KF, Nerlich AG. Classification of age-related changes in lumbar intervertebral discs: 2002 volvo award in basic science. Spine. 2002; 27(23):2631-44.
23. Berlemann U, Gries NC, Moore RJ (1998) The relationship between height, shape and histological changes in early degeneration of the lower lumbar discs. Eur Spine J 7(3):212-217
24. Weiler C, Lopez-Ramos M, Mayer HM, Korge A, Siepe CJ, Wuertz K, Weiler V, Boos N, Nerlich AG (2011) Histological analysis of surgical lumbar intervertebral disc tissue provides evidence for an association between disc degeneration and increased body mass index. BMC Res Notes 4(1):497
25. Corso JJ, Alomari RS, Chaudhary V (2008) Lumbar disc localization and labeling with a probabilistic model on both pixel and object features. In: Proceeding of MICCAI, vol 5241 of LNCS Part 1. Springer, pp 202–210
26. Alomari RS, Corso JJ, Chaudhary V (2009) Abnormality detection in lumbar discs from clinical MR images with a probabilistic model. In: Proceeding of CARS
27. Alomari RS, Corso JJ, Chaudhari V, Dhillon G (2009) Desiccation diagnosis in lumbar discs from clinical MRI with a probabilistic model. In: Proceeding of ISBI'09, pp 546–549
28. Alomari RS, Corso JJ, Chaudhary V, Dhillon G (2009) Computer-aided diagnosis of lumbar disc pathology from clinical lower spine MRI. Int J Comput Assist Radiol Surg 5(3):287–293

29. Watanabe A, Benneker L, Boesch C, Obata T, Anderson S, Watanabe T (2007) Classification of intervertebral disk degeneration with axial T2 mapping. AJR 189(4):936–942
30. da Silva Barreiro M, Marcello H, Rangayyan R (2014) Semiautomatic classification of intervertebral disc degeneration in magnetic resonance images of the spine. In: 5th ISSNII-IEEE Biosignals and biorobotics for better and safer living (BRC) conference, pp 1–5
31. Castro-Mateos I, Hua R, Pozo JM, Lazary A, Frangi AF (2016) Intervertebral disc classification by its degree of degeneration from T2 weighted magnetic resonance images. Eur Spine J 25(9):2721–2727
32. Ronnerberger O, Fischer P, Brox T (2015) U-net: convolution networks for biomedical image segmentation. In: MICCAI, Springer, LNCS, vol 9351, pp 234–241
33. Jamaludin Amir, Kadir Timor, Zisserman Andrew (2017) SpineNet: automated classification and evidence visualization in spinal MRIs. Med Image Anal 41:63–73
34. Kim KI, Kwon Y (2010) Single-image super-resolution using sparse regression and natural image prior. IEEE Trans Pattern Anal Mach Intell 32(6):1127–1133
35. Ojala T, Pietikainen M, Maenpaa T (2002) Multiresolution gray-scale and rotation Invariant texture classification with local binary patterns. IEEE Trans Pattern Anal Mach Intell 24:971–987
36. Krizhevsky A, Sutskever I, Hington GE (2012) ImageNet classification with deep convolution neural networks. In: NIPS, pp 1106–1114
37. Cortes C, Vapnik V (1995) Support-vector networks. Mach Learn 20:273–297

Thermal Image Segmentation of Facial Thermograms Using K-Means Algorithm in Evaluation of Orofacial Pain

Nida Mir, U. Snekhalatha, Mehvish Khan and Yeshi Choden

Abstract The study aims at analyzing skin surface temperature, aided by the thermal camera, a supporting software, and application of k-means algorithm, and feature extraction in MATLAB to diagnose dental diseases, specifically, orofacial pain. The thermal camera is employed for capturing thermal images of the Left, Right, and Front profiles of all the subjects taken into account. MATLAB-based image segmentation using k-means algorithm, and feature extraction was carried out for control and test group data. The results obtained from the study depict that the mean temperature difference of maximum, minimum and average values of temperature recorded were found to be 1.09% in the front, 3.78% in the right, and 3.97% in the left facial regions between the normal subjects and abnormal diseased subjects. Of the regions examined using thermography, and subsequent feature extraction, the right and left sides show almost similar percentage differences, that is, of 3.78 and 3.97%. These findings point toward a clear, and significant rise of temperature, due to presence of infections, or ailments in the From this data it is safe to infer that, presence of infections, significantly increases the temperature of the region they are present in, and hence give an indication of possible application of thermography in dental disease detection.

N. Mir · U. Snekhalatha (✉) · M. Khan · Y. Choden
Department of Biomedical Engineering, SRM University, Kattankulathur, Chennai 603203, Tamilnadu, India
e-mail: sneha_samuma@yahoo.co.in

N. Mir
e-mail: Nidamir33@yahoo.com

M. Khan
e-mail: Meh.pahtan13@gmail.com

Y. Choden
e-mail: wishfangle@gmail.com

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_56

1 Introduction

Orofacial pain is a general term covering any pain usually used to describe symptoms of pain in the mouth, head and neck region. Facial and oral pain, both are included in Orofacial pain. Out of this, pain within the mouth is called orofacial pain, whereas Facial pain includes the pains anterior to ears, below the canthometal line, and over the neck [1]. The most commonly reported symptom of orofacial pains is toothache (57.6%) [2]. Prevalence of OFP was 26.94% which is more common in age group of 35–44 yrs (10.3%), females (14%) and low-income group (9.05%) [3]. These diseases might go unnoticed, or undiagnosed, leading to increased infection or severe problems. Common methods of diagnosis include clinical intra- and extra-oral examination, radiographic method, occlusal examination, and fiber-optic Transillumination, Electrical conductance, quantitative light-induced fluoroscope [4, 5]. Thermography can also be used for studying the physiology of skin temperature of the subjects, and can serve as a non-invasive imaging modality, for detection of dental disorders, according to the amount of heat, and consequently IR rays emitted by the subject's body.

Gratt et al. diagnosed chronic orofacial pain and TMJ disorders by measuring the skin temperature in facial region, using static area telethermography and classified patients with orofacial pain and normal subjects attaining 92% accuracy [6] a few researchers including Durnovo et al., and Pogrel et al. suggested the use of thermography as a potential diagnostic tool for the evaluation of diseases in maxillofacial areas [7, 8].

The aim and objectives of the study were to analyze the skin surface temperature distribution in dental disorders using thermal imaging and to segment the region of interest using k-means algorithm and to perform statistical feature extraction for the total population studied.

2 Methodology

2.1 Study Design and Population

The study was approved by an institutional ethical committee of SRM medical college and Hospital research centre, consisted of a total of 10 normal subjects and 10 subjects suffering from OFP. The institutional scientific committee of SRM medical college and Hospital research centre had approved the study, and the participant's signed the informed consent form.

2.2 Procedure

The subjects were made to sit in a room with the temperature at approximately 20 °C with humidity of 45–50%. Subjects with beard, or wearing ornaments or jewelry within the region of interest that could not be removed were excluded. The distance between the subjects face, and the camera (FLIR A300) was standardized at 1 m, and thermal images of the left, right, and front profiles of the face region of both the subject groups (normal and diseased) were captured. FLIR software was used for image analysis, including measurement of average skin temperature using rectangular and elliptical area tool for ROI selection and further processing was done using MATLAB R2012a (Math Works Inc., Natick, MA, USA).

2.3 Image Segmentation Algorithm

1. Images of left, right, and front profiles of test group and control group subjects were acquired using FLIR 300 thermal camera
    1. images captured were converted from RGB Color Space to $L*a*b*$ Color Space (L is lightness, varying from 0 to 100, '*a*' refers to green-red component and '*b*' refers to blue-yellow component)
    2. the image was then segmented into 16 different clusters to obtain the Hot Spot
    3. region separately, using k-means clustering
    4. images were then labeled according to the results obtained from k-means algorithm.

Table 1 The mean of maximum, minimum and average values of temperature recorded from the front, left and right profiles of the normal and diseased subjects are calculated and depicted in Table 1. A difference of 1.09% in the front, 3.78% in the right and 3.97% in the left facial regions between the normal subjects and abnormal diseased subjects was calculated. Of the regions examined using thermography, and subsequent feature extraction, the right and left sides show almost similar percentage differences, that is, of 3.78 and 3.97% respectively. The average temperature recorded in the region of interest, of front, left and right profiles, of the diseased patients is always higher in comparison to that of normal subjects, though the difference fluctuates, as sides differ.

Table 2: The temperature values collected from front, left and right profiles of the normal and diseased subjects are recorded and are analyzed statistically. The mean, standard deviation, variance, Median, skewness, kurtosis and moment obtained from this analysis are depicted in this table. The values in diseased subjects are present, unlike normal cases, where they are repetitively zero.

Figure 1 shows the right profile of the normal subject. Figure 1a displays the original thermogram to be processed by K-means clustering. Figure 1b, c displays

Table 1 Skin surface temperature measured at various facial regions using thermal imaging

Facial region		Temperature		
		Max	Min	Mean
Front	Normal	35.2875	25.25	31.7125
	Diseased	35.7625	26.35	32.0625
	% difference	1.33	4.26	1.09
Right	Normal	35.0625	25.3375	30.7625
	Diseased	36.1	26.2125	31.95
	% difference	2.91	3.394	3.78
Left	Normal	35.2125	26.1625	30.85
	Diseased	36.0875	26.4125	32.1
	% difference	2.45	0.951	3.97

Table 2 Feature extraction performed for the segmented image of normal and orofacial pain

Feature extracted parameters	Region	Mean	Std. dev	Variance	Median	Skewness	Kurtosis	Moment
Normal ($n = 10$)	Front	0	0	0	0	0	0	0
	Left	0	0	0	0	0	0	0
	Right	0	0	0	0	0	0	0
Orofacial pain ($n = 10$)	Front	56.7201	68.3811	4.69E+03	19.875	1.34949	4.24321	4.69E+03
	Left	56.1394	67.6715	4.59E+03	18.375	1.33023	4.16089	4.59E+03
	Right	57.0711	68.4343	4.70E+03	19.875	1.31145	4.08315	4.70E+03

the color-based segmented images using K-means clustering. Figure 1d displays the segmented image after thresholding. Similarly, Fig. 2 shows the right profile of the diseased subject. Figure 2a displays the original thermogram and Fig. 2b, d displays the various color-based segmented images using K-means clustering. Figure 2c displays the high-temperature regions due to inflammation or infection. Figure 2e displays the segmented image of the highest temperature regions after thresholding. Due to the absence of high-temperature zone in normal subjects, the regions of maximum temperature are not segmented contrary to diseased subjects.

3 Discussion

In the literature, the methods used for diagnosis of dental diseases include observation of symptoms, anatomical changes, and X-rays. This study, aims at establishing

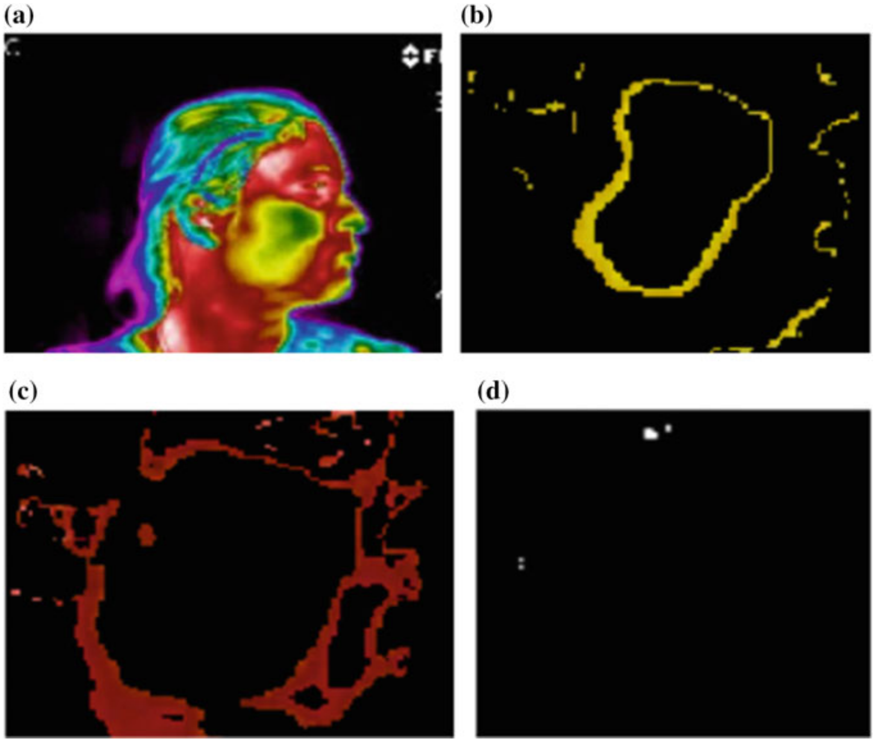


Fig. 1 Normal subject right profile **a** input thermal image, **b** and **c** random clusters obtained using the algorithm, **d** segmented image (absence of high temperature)

thermography as a means, or at least, an aid, in the detection of such diseases, which might otherwise go unnoticed, or might be difficult to rule out, or diagnose. A safe inference, that is, the temperature in case of, diseased patients is higher, can be made, and used as a support for the study. This is specially, advantageous for asymptomatic diseases, as detection of such abnormalities, still remains a matter of concern for the medical professionals.

Gratt et all selected a region of interest, in normal and diseased subjects and classified them into normal, hot and cold, on the basis of temperature recorded in that region [9, 10]. Another study conducted by Dibai Filho et al. suggested that, within the upper trapezius muscle region, infrared imaging was suitable for diagnosis of trigger points in the in clinical and research practices [11].

In a separate study conducted in 2014, by Haddad et al., it was significantly apparent that the temperature measured at anterior temporal muscle regions and masseter muscles in controls were higher than myogenic TMD volunteers, along with posing the suggestion of thermography being a possible tool in diagnosis or aid of other diagnostic methods [12].

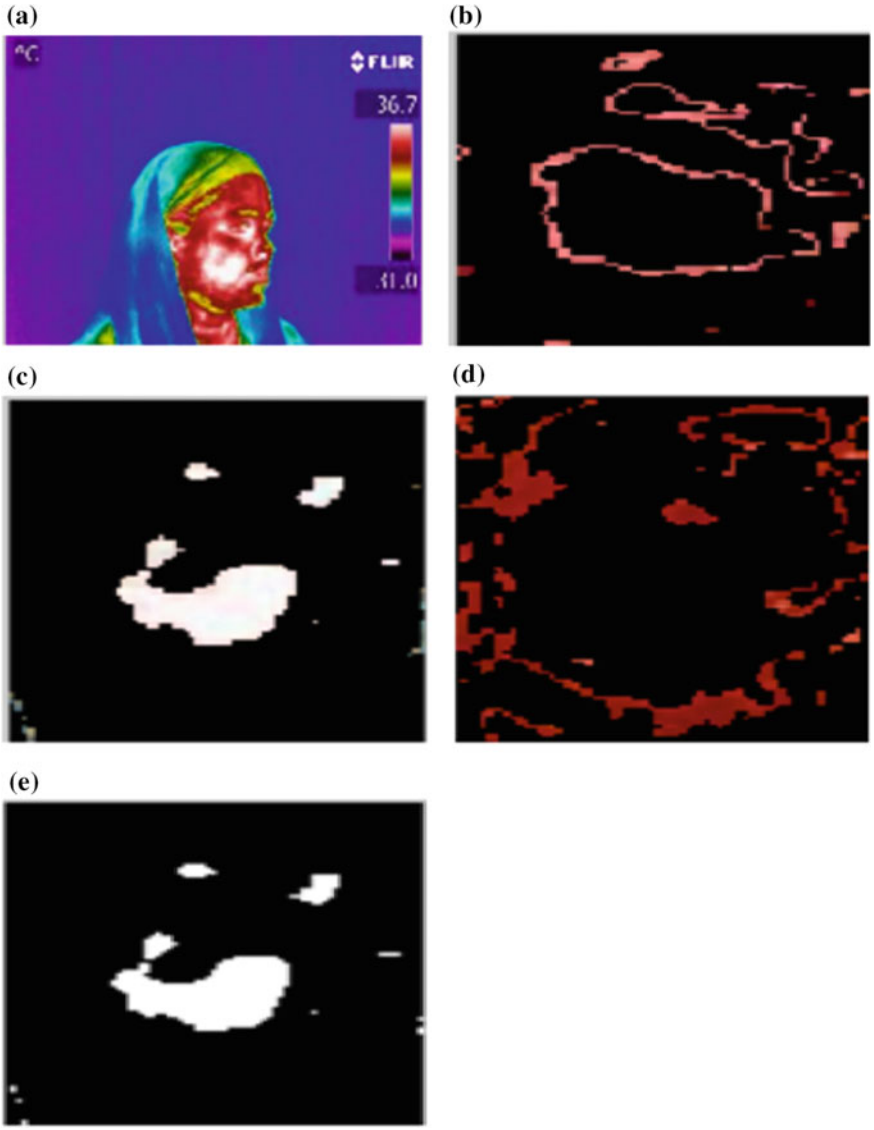


Fig. 2 Diseased subject right profile **a** input thermal image, **b**–**d** random clusters obtained using the algorithm, **e** segmented image showing high temperature alone

Also the studies, like that carried out by McBeth et al., Biagioni et al., etc., are in agreement with the current study, implying the possible use of thermography in future for better diagnosis [13–16].

4 Conclusion

Thermography can be used as a complementary diagnostic tool in dentistry with reference to the facial thermograms as there is a notable escalation of regional temperature. The mean of maximum, minimum, and average values of temperature recorded from the front, left, and right profiles of the normal and diseased subjects were calculated and subsequent feature extraction was implemented. The right and left profiles showed a comparable percentage difference (i.e., 3.78 and 3.97%). As inferred, thermographic findings differ with different subjects' side profiles, however, the average temperature of the diseased subjects are always higher than that of the normal subjects.

References

1. Macfarlane TV, Blinkhorn AS, Davies RM, Kincey J, Worthington HV (2002) Orofacial pain in the community: prevalence and associated impact. Commun Dent Oral Epidemiol 30:52–60
2. Oberoi SS, Hiremath SS, Yashoda R et al (2014) Prevalence of various orofacial pain symptoms and their overall impact on quality of life in a tertiary care hospital in India. J Maxillofacc Oral Surg 13:533. https://doi.org/10.1007/s12663-013-0576-6
3. Manjunath G, Prasad KVV (2012) Prevalence of Orofacial pain among urban adult population of Hubli—Dharwad and its impact on daily. J Indian Assoc Public Health Dent 10:7–18
4. Abesi F, Mirshekar A, Moudi E, Seyedmajidi M, Haghanifar S, Haghighat N, Bijani A (2012) Diagnostic Accuracy of digital and conventional radiography in the detection of non-cavitated approximal dental caries. Iran J Radiol 9(1):17–21
5. Gomez J (2015) Detection and diagnosis of the early caries lesion. BMC Oral Health 15(Suppl 1):S3
6. Gratt BM, Graff-Radford SB, Shetty V, Solberg WK, Sickles EA (1996) A six-year clinical assessment of electronic facial thermography. Dentomaxillofacial Radiol 25:247–255
7. Durnovo EA, Potekhina YP, Marochkina MS, Yanova NA, Sahakyan MY, Ryzhevsky DV (2014) Diagnostic capabilities of Infrared thermography in the examination of patients with diseases of maxillofacial area. Clin Med 6(2):61–65
8. Pogrel MA, Yen CK, Taylor RC (1989) Infrared thermography in oral and maxilla facial surgery. Oral Surg Oral Med Oral Pathol 67(2):126–131
9. Gratt BM, Sickles EA, Ross JB (1994) Thermographic characterization of an internal derangement of the temporomandibular joint. J Orofacial Pain 8:197–206
10. Gratt BM, Sickles EA, Wexler CA (1993) Thermographic characterization of osteoarthrosis of the temporomandibular joint. J Orofacial Pain 7:345–353
11. Dibai-Filho AV, Guirro EC, Ferreira VT, Brandino HE, Vaz MM, Guirro RR (2015) Reliability of different methodologies of infrared image analysis of myofascial trigger points in the upper trapezius muscle. Braz J Phys Ther 19:122–128
12. Haddad DS, Brioschi ML, Vardasca R, Weber M, Crosato EM, Arista ES (2014) Thermographic characterization of masticatory muscle regions in volunteers with and without myogenous temporomandibular disorder: preliminary results. Dentomaxillofacial Radiol 43:20130440
13. Gratt BM, Sickles EA (1995) Electronic facial thermography: an analysis of asymptomatic adult subjects. J Orofacial Pain 9:255–265
14. Biagioni PA, Longmore RB, McGimpsey JG, Lamey PJ (1996) Infrared thermography. Its role in dental research with particular reference to craniomandibular disorders. Dentomaxillofacial Radiol 25:119–124

15. Komoriyama M, Nomoto R, Tanaka R, Hosoya N, Gomi K, Iino F, Yashima A, Takayama Y, Tsuruta M, Tokiwa H, Kawasaki K, Arai T, Hosoi T, Hirashita A, Hirano S (2003) Application of Thermography in dentistry-visualization of temperature distribution on oral tissues. Dent Mater J 22(4):1–7
16. McBeth SA, Gratt BM (1996) A cross-sectional thermographic assessment of TMJ problems in orthodontic patients. Am J Orthod Dentofac Orthop 109:481–488

Analysis of Web Workload on QoS to Assist Capacity

K. Abirami, N. Harini, P. S. Vaidhyesh and Priyanka Kumar

Abstract Workload characterization is a well-established discipline, which finds its applications in performance evaluation of modern Internet services. With the high degree of popularity of the Internet, there is a huge variation in the intensity of workload and this opens up new challenging performance issues to be addressed. Internet Services are subject to huge variations in demand, with bursts coinciding with the times that the service has the most value. Apart from these flash crowds, sites are also subject to denial-of-service (DoS) attacks that can knock a service out of commission. The paper aims to study the effect of various workload distributions with the service architecture 'thread-per-connection' in use as a basis. The source model is structured as a sequence of activities with equal execution time requirement with an additional load time of page (loading embedded objects, images, etc.). The threads are allocated to the requests in the queue; leftover requests if any are denied service. The rejection rate is used as a criterion for evaluation of the performance of the system with a given capacity. The proposed model could form a basis for various system models to be integrated into the system and get its performance metrics (i.e. QoS) evaluated.

1 Introduction

Today Internet has emerged as the default platform for application development. Unfortunately, modern applications demand more complexity than traditional applications. As the Internet was not designed to suit the requirements of modern applications, the execution results in high frustration of users. This factor demands a research on how the existing infrastructure could be modified for efficient execution of modern web applications [1]. The complexity exhibited by applications are multifold (process, data, load, configuration, scale, etc.). With an intent to improve the

K. Abirami (✉) · N. Harini · P. S. Vaidhyesh · P. Kumar
Department of Computer Science and Engineering, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Coimbatore, India
e-mail: k_abirami@cb.amrita.edu

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_57

performance of web applications a study on workload characterization is compelling. Many researchers have focused their study on understanding the characteristics and intensity of workloads. In this work, we discuss the role of workload models for resource assignment in the scenario of the e-commerce application. The impact of the workload on system properties and behavior is analyzed using a capacity planning model. The proposed system evaluates the Quality of Service (Qos) and Quality of Experience (QoE) perceived by the users for different workload distributions. These observations could aid in framing security mechanisms, recommendation engines, data distribution policies, etc.

When the system is scaled, the work also presents major findings from experimentation indicating performance implications. The rest of the paper is organized as follows: Sect. 2 presents a comprehensive overview of the literature on different workload distributions. Section 3 summarizes the characterization methodologies and related measurement process. Section 4 presents the results and analysis of experimentation and finally, Sect. 5 presents concluding remarks.

2 Literature Review

Many research work addresses the black box approach for the assessment of performance based on workloads. Rejection rates have a huge impact on the performance of the system [2]. Recent rates have a huge impact on the performance of the system [2]. Recent studies have also considered performance measurements based on user behavior patterns and businesses [3, 4]. The response time metric has been chosen in most of the research work for performance evaluation [5]. The Zipf law's applicability of web workloads is addressed by Levene et al. [6] and Menasce et al. [7]. Mi et al. [8] and Harini and Padmanabhan [9] discuss the need for stationary of arrival processes to study and characterize web load. Harini and Padmanabhan [9, 10] addresses the issue of the presence of malicious request in the incoming lot which needs to be weeded out before the commencement of processing. Workload management is a process of effective workload distribution to achieve optimal performance and productivity levels. Modelling workload distributions would aid one to understand the performance and the scalability of the system. Workload model of an application depicts how the application would perform in the given infrastructure. The performance is usually assessed using Service-Level Agreements (SLA). Little Theorem gives a relationship between the average number of users, arrival rate and average time, an end user spends in the system.

The theorem state that

$$L = \lambda N$$

where λ is the arrival rate and L is the effective arrival rate.

The only prerequisite being system should not preempt and must be stable. The arrival pattern of the request can be modeled based on different probability distributions like exponential, normal, binomial, Poisson, Zipfian.

2.1 Distributions

2.1.1 Exponential Distributions

Exponential distribution is a well-known concept in the theory of probability and statics. The distribution denotes the time between two events in processes where the events are continuous and occur independently. The key property of the distribution is memorylessness. This general exponential distribution is given by

$$f(x;\lambda) = \left\{ \begin{array}{cc} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{array} \right\}$$

where λ greater than 0 is the rate parameter. The distribution is well supported in the interval 0 to infinity. This distribution is mainly used to model service times rather than arrival patterns. These can have a strong effect on performance evaluation results.

2.1.2 Normal Distribution

Normal distribution is a very commonly used distribution to determine whether an observation falls between two extreme limits. This distribution is used to model random variables in natural and social sciences. The normal distribution is used in real-valued random variables, where the distributions are not available. The general normal distribution is given by

$$f\left(x|\mu, \sigma^2\right) = \frac{1}{\sigma}\psi\left(\frac{x-\mu}{\sigma}\right)$$

The standard normal deviate is given by Z where

$$Z = \frac{(X-\mu)}{\sigma}$$

These could be used to model the peak of arrivals and at a more concrete level, it can help one to identify results of random effects on workloads.

2.1.3 Poisson Distribution

The application of Poisson distribution in traffic problems is not new. A Poisson distribution is a probability distribution of a discrete random variable that represents the number of statistically independent events occurring within a unit of time or space. Time-based Poisson variables are more popular. Given the expected Value μ of the Poisson variable x the probability function is defined as the probability of observing k events in an interval is given by the equation

$$P(k \text{ events in interval}) = e^{-\lambda}\frac{\lambda^k}{k!}$$

where the average number of events per interval e is the number 2.71828 … (Euler's number) the base of the natural logarithms k is any natural number, $k! = k$ $(k\ 1)$ $(k\ 2)\ 2\ 1$ is the factorial of k. This could be used to model the rate of arrival of request patterns and it could be also used to measure the performance when requests are queued in the system.

2.1.4 Zipf Distribution

A Zipf distribution is sometimes referred to as zeta distribution. This is particularly used for modeling rare events. The probability density function for Zipf distribution is the nth raw moment is defined as the expected value of X_n:

$$m_n = E\left(X^n\right) = \frac{1}{\zeta(s)}\sum_{k=1}^{\infty}\frac{1}{k^{s-n}}.$$

The series on the right is just a series representation of the Riemann zeta function, but it only converges for values of *s-n* that are greater than unity. Thus:

$$m_n = \begin{cases} \frac{\zeta(s-n)}{\zeta(s)} & \text{for } n < s - 1 \\ \infty & \text{for } x \geq 0 \end{cases}$$

Note that the ratio of the zeta functions is well defined, even for $n > s-1$ because the series representation of the zeta function can be analytically continued. The Zipf distribution turns out to better describe varied human activities. It is a good model for popularity distribution. This does not change the fact that the moments are specified by the series itself, and are therefore undefined for large n.

2.1.5 Binomial Distribution

The Binomial distributions will have two outcomes, success or failure. The experiment can have n number of trials and the outcomes are independent. The general equation of the distribution is given by the following:

$$b(x; n, p) = \binom{n}{r} p^x (1-p)^x$$

where,

n represents the number of trials, x represents the number of successes, p represents the probability of success in an individual trial.

The distribution could be used for modeling random arrival patterns, study effects of peak load, perform resource assignments, etc.

2.2 Summary of Findings

A special case of performance evaluation that deserves individual attention is capacity planning. Many research works propose different methodologies for setting up configurations that would provide desired performance. The required system capacity obviously depends on workload intensity, i.e., one needs more capacity to do more work. The relationship between capacity and workload is often not linear. Researchers have also stated that burst is an important attribute that contribute to capacity planning. Burst refers to large fluctuations in workload intensity.

A good characterization technique thus requires a clear understanding of burst characteristics. A combination of system model with workload characterization can enhance the performance of the system. Although individual schemes specific to Internet services have been explored in large, to the best of our knowledge a comprehensive study based on multiple workload distribution based analysis with its performance assessment has not been addressed to a greater extent.

2.3 Problem Statement

To build a system capable of characterizing the performance of a system model for varied workload distributions, which could aid in capacity planning, arriving at optimal configuration to improve QoS, assist in data movement with applied security features.

3 Proposed System

Though Internet services remain simple in the structure at the start they become more complex when functionalities of the service expand. The block diagram for service architecture used for experimentation is shown in Fig. 1.

3.1 Model Description

The basic service is taken as composed of n sequential activities. A single processor system with T threads is used as a basis for service. The service request is taken as a stationary random process having a selected distribution (Normal, Binomial, Exponential, Poisson, Zipf) with its associated parameters.

3.1.1 Request Arrival Pattern

A Service request is taken as a stationary random process with an associated mean and variance attributes. The proposed system considers discrete random arrival pattern. Each arrival is independent of the previous arrival. An arrival set is characterized with a number of incoming requests and a class type associated with it. The arrival capacity is not limited.

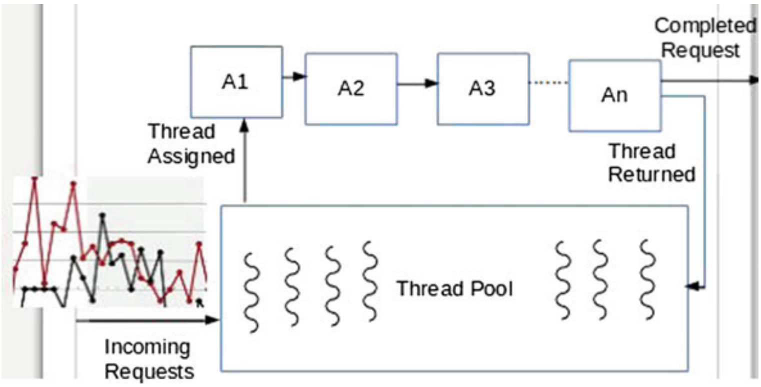


Fig. 1 Service architecture

3.1.2 Request Characteristics

A request is characterized by arrival time and service class. Each request is assumed to have the same number of tasks to be completed. Each request is assumed to have a load time that is dependent on the dynamic content of the webpage and an associated service time.

3.1.3 Service Scheme

At the beginning of every time slots, threads from the request that completed execution are returned to the free pool of threads. The incoming requests are assigned threads for until request list is exhausted or threads in the thread pool are exhausted. Once a thread is allocated to a request it remains associated with the request until completion of execution. There can be time slots when threads are free and those when request are dropped. Both do not happen simultaneously. An extensive simulation was carried out with different distributions. The variations in terms of request drops are modeled and presented in Sect. 4.

3.1.4 Service Scheme Algorithm

The algorithm for modeling arrival:

Step 1: Generate the total number of arrivals for each time slot based on the random number generated by the distribution parameters.
Step 2: Assign the arrivals randomly to n categories as category1, category2 …category$n$ with:

$$\sum_{k=1}^{n}(\text{category1} + \text{category2} + \cdots \text{category}n)$$

It should be equal to a number of arrivals in the time slot category1 = rand()% number of arrivals.

3.1.5 Resource Allocation

Step 1: Initialize index and rejections as 0, Initialize T as maximum number of threads
Step 2: For the arrivals in timeslot allocate thread from the thread pool
Step 3: Update thread counter in the thread pool
Step 4: If not enough threads for allocation.

```
for index in value:
        if value[index]≥ rejection threshold:
                rejection threshold –=value[index]
        else
                rejections +=value[index]
return rejections
```

4 Results and Discussions

The effect of workloads based on different distributions was studied through extensive simulation process. The representative results for thread pool capacity 100 are presented in Fig. 2. The simulation run duration in each case was selected in such a way that all possible service request values appeared enough number of times to bring out all behavioral characteristics. The rejection of requests for different distributions is also presented in Fig. 2. To facilitate service differentiation, three categories of arrivals were considered (this could be used for priority scheduling). Automated service history collection which enables culling intelligent information out of it is also collected by the system.

4.1 Measuring Overhead in Dynamic Pages

Processing dynamic webpage requires additional page load time that includes loading time of images, audio video links etc., To understand the effect of this additional time a webpage with following specification (i.e.,) load time approximately 9.14 s etc., is presented in the table. Image loading time for a website is approximately 1–2 ms. As an example, the website "www.amrita.edu", the total no of requests is 151. The size of the page is 2.6 MB, load time is 9.14 s. Out of this 151 requests 94 (64.1%) requests are given for image. The download time for the images is approximately 1–2 ms each.

This paper clearly forms a useful contribution for assessing the impact of load on the web server for a selected system configuration. Our observations indicated that system model integration with the architecture can enable one to analyze the performance of an Internet service. High rejection rates indicate the need for an increase in the capacity of the system. While doing so, one should ensure that the resources are not underutilized. A low percentage of rejection rates with Zipf distribution indicate the identical behavior of incoming requests. The rejection rate under the Normal distribution shows the random occurrences of peak load in the traffic. The rejection rate under Binomial and Poisson distribution clearly demand scaling the system capacity.

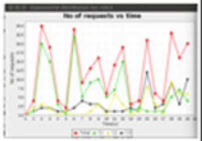
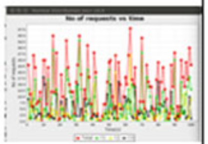
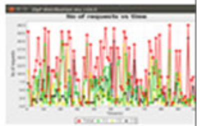
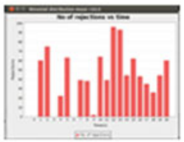
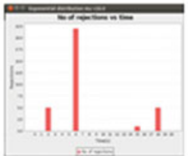
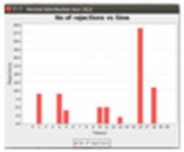
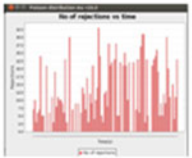
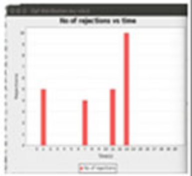
System Capacity/ No.of Threads	100			100			100			100			100		
Mean	26			26			26			26			26		
variance	2			2			2			2			2		
Distribution	Binomial			Exponential			Normal			poisson			Zipf		
Generated Traffic/ Workload															
Rejection/Timeslot															
	C1	C2	C3	C1	C2	C3	C1	C2	C3	C1	C2	C3	C1	C2	C3
%of traffic	48.43	22.86	28.70	51.70	20.06	28.23	49.82	20	30.17	46.91	25.30	27.77	44.68	22.97	32.3
%of rejection	56.03	20.78	23.17	61.6	22.4	16	62.8	17.7	19.4	56.1	22.2	21.6	27.0	10.5	62.3
total rejection			76.58			9.91			25.85			45.79			7.72

Fig. 2 Table 1

5 Concluding Remarks

Modern web services have thrown up many unconventional challenges for monitoring QoS. Although ways for monitoring QoS parameters have been addressed extensively in the literature, the methodologies and techniques applied for creating workload models are strictly related to the objectives of the studies. With the aim of studying the effect of the movement of data in distributed systems in terms of response metric, the scheme proposed in the paper was implemented. The scheme enabled to understand the impact of different distributions on system performance (Completed vs. Rejected Services). Experimentation clearly revealed the effect of dynamic contracts in processing concurrent requests. Schemes like loading essential partial images rather than complete contents could be used to improve the QoS.

References

1. Calzarossa MC, Massari L, Tessera D (2016) Workload characterization: a survey. ACM Comput Surv (CSUR) 48(3):48
2. Galletta DF, Henry R, Mccoy S, Polak P (2004) Web site delays: how tolerant are users. J Assoc Inform Syst pp 1–28
3. Goncalves MA, Almeida JM, dos Santos LG, Laender AH, Almeida V (2010) On popularity in the blogosphere. IEEE Internet Comput 14(3):42–49
4. Gusella R (1991) Characterizing the variability of arrival processes with indexes of dispersion. IEEE J Sel Areas Commun 9(2):203–211
5. Gunther NJ (2001) Performance and scalability models for a hypergrowth e-commerce web site. In: Performance engineering, state of the art and current trends. Springer, London, UK, pp 267–282
6. Levene M, Borges J, Loizou G (2001) Zipfs law for web surfers. Knowl Inf Syst 3(1):120–129
7. Menasce D, Almeida V, Riedi R, Ribeiro F, Fonseca R, Meira W Jr (2000) In search of invariants for e-business workloads. In: EC 00: proceedings of the 2nd ACM conference on electronic commerce, New York, NY, USA. ACM, pp. 56–65
8. Mi N, Casale G, Cherkasova L, Smirni E (2008) Burstiness in multi-tier applications: symptoms, causes, and new models. In: Middleware 08Proceedings of the 9th ACM/IFIP/USENIX international conference on middleware, New York, NY, USA. Springer, New York, Inc., pp 265–286
9. Harini N, Padmanabhan TR (2012) A secured-concurrent available architecture for improving performance of web services. In: Communications in computer and information science, vol 292, no 1. Springer, pp 621–631
10. Harini N, Padmanabhan TR (2013) Admission control and request scheduling for secured-concurrent-available architecture. Int J Comput Appl 63(6):24–30

Content-Based Image Retrieval Using Hybrid Feature Extraction Techniques

B. Akshaya, S. Sruthi Sri, A. Niranjana Sathish, K. Shobika, R. Karthika and Latha Parameswaran

Abstract Images consist of visual components such as color, shape, and texture. These components stand as the primary basis with which images are distinguished. A content-based image retrieval system extracts these primary features of an image and checks the similarity of the extracted features with those of the image given by the user. A group of images similar to the query image fed is obtained as a result. This paper proposes a new methodology for image retrieval using the local descriptors of an image in combination with one another. HSV histogram, Color moments, Color auto correlogram, Histogram of Oriented Gradients, and Wavelet transform are used to form the feature descriptor. In this work, it is found that a combination of all these features produces promising results that supersede previous research. Supervised learning algorithm, SVM is used for classification of the images. Wang dataset is used to evaluate the proposed system.

B. Akshaya · S. Sruthi Sri · A. Niranjana Sathish · K. Shobika · R. Karthika (✉)
Department of Electronics and Communication Engineering, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Coimbatore, India
e-mail: r_karthika@cb.amrita.edu

B. Akshaya
e-mail: akshayabalan96@gmail.com

S. Sruthi Sri
e-mail: shruthisridhar.10@gmail.com

A. Niranjana Sathish
e-mail: niranjana.sathish97@gmail.com

K. Shobika
e-mail: shobi2197@gmail.com

L. Parameswaran
Department of Computer Science and Engineering, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Coimbatore, India
e-mail: p_latha@cb.amrita.edu

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_58

1 Introduction

With a splurge of visual data, it is a vastly cumbersome task to scan tens of thousands of images manually. The content-based image retrieval system extracts the basic features of every image in a dataset and compares the same with those of the image provided by the user. General flow states that post the query matching, the system ranks the images in a descending order of similarity with the given input query and the output is all the images that are ranked highest. Every image will have three basic components: Shape, Color, and Texture. A CBIR relies on these extracted features, individually or combinations of them, to extract and run similarity algorithms on them.

While CBIR completely relies on the contents of the image itself, image retrieval using the metadata of images can also be done. There is textual data associated with each image and traditional methods of retrieval, such as retrieval using keywords, can be done. While the annotation process is time consuming and laborious, there is also an additional limitation with this system which pertains to lack of standardization. Meaning, no two users perceive an image in the same way. Since there is no standard way to perceive an image, multiple users will give multiple annotations which are not likely to match. This is will also result in large amounts of junk data which is undesirable. Thus, CBIR is more suitable for large amounts of visual data.

In this paper, a combination of color, texture, and shape features are extracted in order to maximize the accuracy of extraction and efficacy of output. A proposed method uses best of multiple techniques to improve quality of output and reduce error margins.

In order to enhance the results of a robust dataset (Wang dataset), the images are classified by Supervised Vector Machines (SVM) algorithm. SVM is the most efficient supervised learning algorithm for image recognition, face recognition, speech recognition, and face detection. It is reliable, accurate and is most efficient for binary classification.

2 Literature Survey

The properties of Hue, Saturation, and Intensity values color space is analyzed in [1]. The values are varied and the visual perception is studied. The saturation value is used to decide if the Hue or the Intensity of the pixel is closer to human perception.

Various CBIR tools are compared in [2]. From this comparison, it is observed that most of the systems use color and texture features. Shape feature is not as common. Layout feature is very rarely used. Retrieval techniques based on a single feature worked well only for a specific set of images.

One of the most commonly used color feature in CBIR system is color histogram, [3–5]. Color Histogram concentrates only on the proportion of the number of various types of colors in an image, but does not focus on the spatial location of the colors.

Noise is not handled efficiently by histograms because they are very sparse. To overcome the drawback of this feature, features such as color-correlogram and color moments are applied. Preprocessing the images will increase the accuracy.

In [6], it is observed that color and texture features are used. Support Vector Machine (SVM) and Euclidean distance are applied to retrieve similar images.

Different approaches of different combinations of color, shape and texture retrieval are compared in [7]. When color (color histogram) and texture features (standard wavelet) are combined, accuracy was enhanced but the feature set was inadequate. On merging color, texture and shape feature (Color moment, Gabor filter, Gradient Vector Flow), the strong feature set was created.

Ecosembles is formed by concatenating different combinations of weak feature views (color, shape, texture, etc.), which must be extracted from images. Histograms are extracted from a group of images with varying numbers of bins for each histogram. By examining the Intelligence, Surveillance, and Reconnaissance data, it is understood that Ecosembles performed slightly better than GIST descriptors coupled with SVM. This is observed in [8].

In [9], images are retrieved separately using the features like color Histogram, Gabor and wavelet transform for texture, and Shape information from Phase congruency (edge detection for any change in illumination and contrast in the image). A combination of these produced an accuracy of 96.4%.

A system for biometric security for CBIR is developed based on the extraction of Shape (moment invariant), Color (Histogram), and Texture (Gabor wavelet) in [10].

The color and texture features are concatenated where Wavelet-Based Color Histogram (WBCH) method is used. The precision of this proposed method is found to be better. The computational steps are reduced with the help of wavelet transform, thereby increasing the retrieval speed [11].

In [12], a relative study on several features like merged color histogram and Gabor transform is performed.

With only statistical entities of the first order such as mean and standard deviation, Gabor wavelet showed better classification results. This method is proven to be slightly superior to the co-occurrence matrix which is usually used for texture classification [13].

A CBIR system in which the features like HOG, SIFT, SURF, and color histogram are used to extract the features of the image and formed a collection of local feature vectors is observed in [14].

3 Proposed Work

The primary goal of the proposed work is to retrieve similar images from a database as a response to passing a single query image. Low-level features of images are used in this approach. The prime motive of this work is to obtain an efficient and less complex image retrieval system. To achieve this, a comparative study on image retrieval using different features is performed. Initially, images are retrieved with

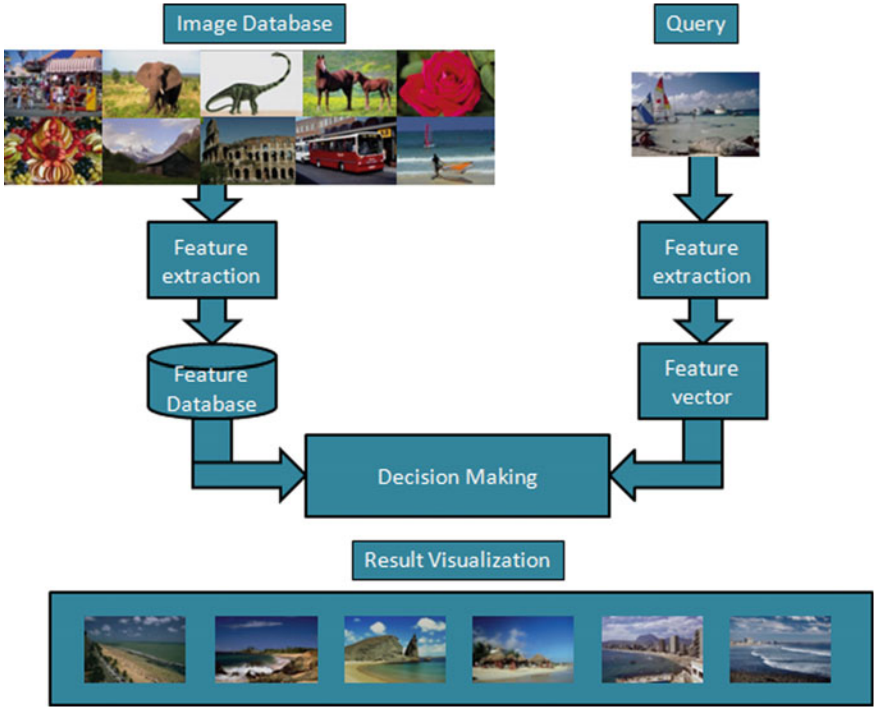


Fig. 1 CBIR system architecture

color, shape and texture features separately. Further, all three features are combined and results are compared.

In this approach, Support Vector Machine (SVM) algorithm is used for classification. The dataset used here is Wang dataset [15], which consists of 1000 images consisting of 10 different classes. Each class has 100 images. The classes include Africa, Rose, Beach, Monument, Bus, Horse, Food, Dinosaur, Scenery, and Elephant. The sizes of the images are either 384×256 or 256×384.

The schematic representation of the CBIR is shown in Fig. 1.

4 Feature Extraction

This is the most crucial stage in CBIR. The feature of each image is extracted and stored in a vector. The process is as described below.

4.1 Color Feature

HSV Histogram. Color Histogram is a schematic impression of the distribution of colors in an image. Although simple and straightforward, Color Histogram is sensitive to change in brightness and does not account for spatial information. Hence, images are converted from Red, Blue, and Green (RGB) toHue, Saturation, Value (HSV) color space. The histogram is computed for the HSV color model as it has an advantage that it separates the chrominance and luminance of an image. Each HSV component is quantized into (8 * 2 * 2) bins which gives a vector of 32 dimensions.

Color Auto Correlogram. A color auto correlogram, which is an indexed table of color pairs and their probabilities, includes spatial correlation of colors and is easy to compute. In color auto correlogram, color distribution is calculated as a function of the distance between two pixels. The image is quantized into 4 * 4 * 4 bins, which give a feature vector of 64 dimensions.

Color Moment. Color moment is the simplistic distribution of color in the image. Likeliness of two images can be compared with the help of color moments. The mean and standard deviation moments are calculated. Mean is the average color of the image and the standard deviation is calculated by taking the square root of the variance. Thus, the first two moments of each channel (RGB) is extracted to form a vector of six dimensions.

In order to compensate for the deficits of each technique, a combination of three techniques are used in order to leverage their best outcomes for color feature extraction.

Mean:

$$E_i = \sum_{j=1}^{N} \frac{1}{N} P_{ij} \tag{1}$$

N Number of pixels in the image
P_{ij} value of the jth pixel in the image at ith color channel.

Standard deviation:

$$\sigma = \sqrt{\left(\frac{1}{N}\sum_{j=1}^{N}\left(P_{ij} - E_i\right)^\wedge 2\right)} \tag{2}$$

E_i mean value for ith color channel of the image.

4.2 Shape Feature

Wavelets are a more general way to represent and analyze multiresolution images. Thus, shape feature extraction is accomplished using wavelet transform and Histogram of Oriented Gradients.

Histogram of Oriented Gradients (HOG). The HOG technique counts the number of occurrences of a specific orientation in each part of the image. It forms a feature vector of 1-N length, where N is the length of the HOG feature. The mean is used to form a feature vector.

Wavelet Transform. The wavelet transform is a multiresolution filtering technique that eliminates noise efficiently. The DWT (2 discrete wavelet transform) is used for detection of edges. Coiflet wavelet is applied with a 3 level decomposition and the mean and standard deviation is used to form a feature vector of 40 dimensions.

4.3 Texture Feature

Gabor Wavelet. Gabor Wavelet is used for the extraction of texture feature. The Gabor representation minimizes uncertainty in space and frequency dimensions and the micro-features extracted characterize texture information. Gabor wavelet filters are applied to each image spanning across four scales and six orientations. This produces a vector of 48 dimensions.

All features from the aforementioned steps are concatenated to form a feature vector of 192 dimensions.

A query image is an input for feature extraction and the feature vector is stored.

5 Classifier

5.1 Support Vector Machines (SVM)

Post the feature extraction process, all the images in the database are classified using SVM. It is a supervised learning algorithm that is used for classification and regression analysis. The approach used is “one-versus-one”, where $\frac{n!}{(n-k)!k!}$ binary classifiers have to be trained for a k-way problem. It differentiates the samples of a pair of classes at a time. When a query image is given as input, a voting scheme is applied to all $\frac{n!}{(n-k)!k!}$ classifiers. The predicted output by the classifier is the class that gets the highest number of ‘+1’ predictions.

6 Result Analysis

Various experiments were performed to show the efficiency of the proposed method. The system receives a single query image and returns 20 similar images from the database. The result is tested using test images from each class. Of 100 images in each class, 95 images are trained and 5 images are used for testing. For each query image, there exist 95 relevant images.

Performance evaluation can be done using numerous metrics. In this paper, precision and recall have been used for performance evaluation.

Precision

Precision is the ratio of retrieved relevant images to the total number of images retrieved.

$$\text{Precision} = \frac{\text{no. of relevant images retrieved}}{\text{total no. of images retrieved}} \tag{3}$$

Recall

Recall is the measure of how many number of truly relevant results are retrieved. A high recall implies that the algorithm has returned most of the relevant images.

$$\text{Recall} = \frac{\text{no. of relevant images retrieved}}{\text{no. of relevant images in the database}} \tag{4}$$

Table 1 shows the average values of the precision and recall of Wang dataset. It shows the retrieval performance when color, texture, and shape features are used in isolation. The average value is taken for each class. Table 2 shows the values based on a combination of color, texture, and shape feature.

Table 3 shows a comparison of the existing technique to the proposed technique. It can be seen that the color feature is more effective compared to shape and texture when the features are used in isolation. It is clear that the proposed system where the features are combined offers the highest value of precision and recall (Figs. 2, 3, 4 and 5).

The results demonstrate that the output obtained by using color, shape and texture features separately do not match with the query image whereas a combination of the three produced much accurate results. The performance measure was calculated based on precision and recall.

7 Conclusion

The goal of this paper is to retrieve images from a database with reliable accuracy by using multiple techniques in tandem with one another. The proposed work introduces an integrated approach to CBIR which helps retrieve similar images from a

Table 1 Retrieval result using color, shape, and texture separately

Class	Color		Shape		Texture	
	Precision	Recall	Precision	Recall	Precision	Recall
Africa	0.8	0.16	0.4	0.08	0.6	0.12
Beach	0.8	0.16	0.6	0.12	0.6	0.12
Monument	0.8	0.16	0.4	0.08	1	0.2
Bus	0.8	0.16	0.6	0.12	0.6	0.12
Dinosaur	0.8	0.16	1	0.2	1	0.2
Elephant	0.8	0.16	0.8	0.16	0.2	0.04
Rose	1	0.2	0.8	0.16	1	0.2
Horse	1	0.2	0.8	0.16	0.4	0.08
Mountain	0.8	0.16	0.6	0.12	0.6	0.12
Food	0.6	0.12	0.4	0.08	0.6	0.12
Mean	0.82	0.164	0.64	0.128	0.66	0.132

Table 2 Retrieval result combining color, shape, and texture

Class	Precision	Recall
Africa	0.8	0.16
Beach	0.8	0.16
Monument	0.8	0.16
Bus	1	0.21
Dinosaur	1	0.21
Elephant	0.8	0.16
Rose	1	0.21
Horse	1	0.21
Mountain	0.8	0.16
Food	0.8	0.16

Table 3 Comparison of the existing and proposed method

	Color	Shape	Texture	Color, shape and texture
Precision	0.82	0.64	0.66	0.88
Recall	0.164	0.128	0.132	0.18

Fig. 2 Precision and recall using color

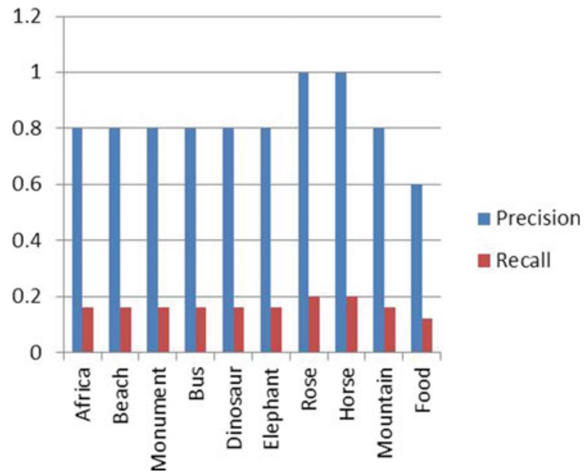
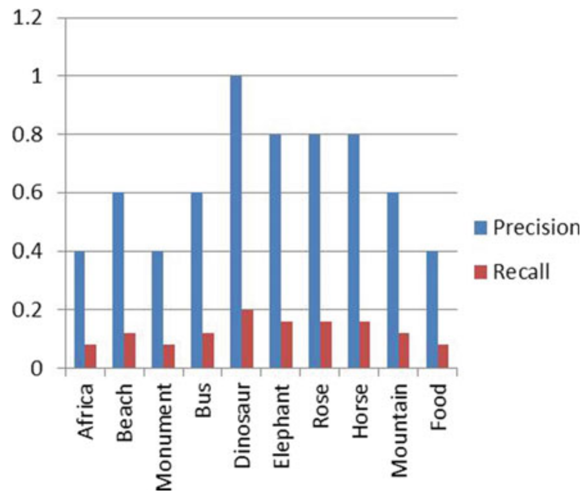


Fig. 3 Precision and recall using shape



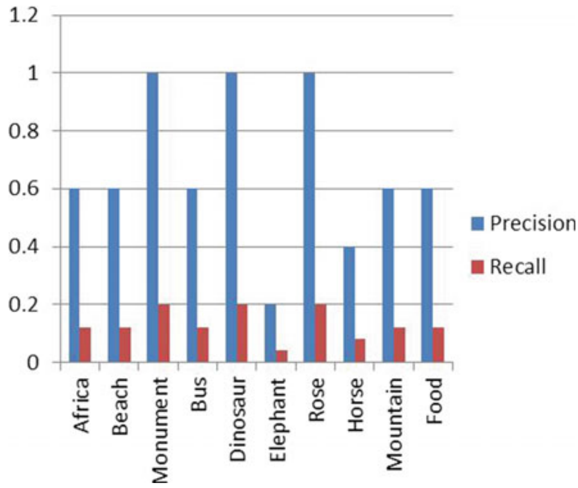


Fig. 4 Precision and recall using texture

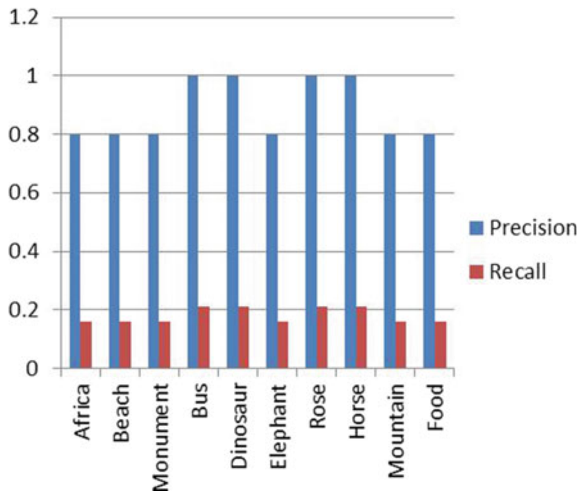


Fig. 5 Precision and recall combining all features

database using SVM. A comparative study on retrieval results was performed and the results using color, shape, and texture features in tandem with one another has given the precision value of 0.88 and recall value of 0.18 for the Wang dataset, which superseded previous research on CBIR. Future refinement of the work will involve research using a bag of words as a feature for larger datasets.

References

1. Sural S, Qian G, Pramanik S (2002) Segmentation and histogram generation using the HSV color space for image retrieval. In: IEEE international conference on image processing
2. Shirazi S, Khan NUA, Umar AI, Razzak MI, Naz S, AlHaqbani B (2016) Content-based image retrieval using texture color shape and region. Int J Adv Comput Sci Appl (IJACSA) 7(1):418–426
3. Iqbal Q, Aggarwal JK (2002) CIRES: a system for content-based retrieval in digital image libraries. In: Seventh international conference on control, automation, robotics, and vision (ICARCV), Singapore
4. Gonde AB, Maheshwari RP, Balasubramanian R (2013) Modified curvelet transform with vocabulary tree for content based image retrieval. Digit Signal Proc 23(1):142–150
5. Manjunath BS, Ohm JR, Vasudevan VV, Yamada A (2001) Color and texture descriptors. IEEE Trans Circ Syst Video Technol 11(6):703–715
6. Giri A, Meena YK (2014) Content based image retrieval using integration of color and texture features. Int J Adv Res Comput Eng Technol (IJARCET) 3(4)
7. Pandey D, Shivpratapkushwah A (2016) Review on CBIR with its advantages and disadvantages for low-level features. Int J Comput Sci Eng 4(7):161–167
8. Rosebrock A, Oates T, Caban J (2013) Ecosembles: a rapidly deployable image classification system using feature-views. In: 12th International conference on machine learning and applications
9. Prakash KSS, Sundaram RMD (2007) Combining novel features for content based image retrieval. In: EURASIP conference focused on speech and image processing
10. Aravind G, Andan HM, Singh T, Joseph G (2015) Development of biometric security system using CBIR and EER. In: IEEE international conference on communication and signal processing (ICCSP)
11. Singha M, Hemachandran K (2012) Content based image retrieval using color and texture. Signal Image Process Int J (SIPIJ) 3(1):39
12. Raghupathi G, Anand RS, Dewal ML (2010) Color and texture features for content based image retrieval. In: Second International conference on multimedia and content based image retrieval
13. Arivazhagan S, Ganesan L, Priyal SP (2006) Texture classification using Gabor wavelets based rotation invariant features. Pattern Recogn Lett 27(16):1976–1982
14. Bagyammal T, Parameswaran L (2015) Context based image retrieval using image features. Int J Adv Inf Eng Technol (IJAIET) 9(9)
15. Li J, Wang JZ (2003) Automatic linguistic indexing of pictures by a statistical modeling approach. IEEE Trans Pattern Anal Mach Intell 25(9):1075–1088. http://wang.ist.psu.edu/docs/related/

Review of Feature Extraction and Matching Methods for Drone Image Stitching

M. Dhana Lakshmi, P. Mirunalini, R. Priyadharsini and T. T. Mirnalinee

Abstract Image stitching is the process of combining multiple overlapping images of different views to produce a high-resolution image. The aerial perspective or top view of the terrestial scenes will not be available in the generic 2D images captured by optical cameras. Thus, stitching using 2D images will result in lack of information in top view. UAV (Unmanned Aerial Vehicle) captured drone images tend to have the high aerial perspective, 50–80% of overlapping of information between the images with full information about the scene. This work comprises of discussion about methods such as feature extraction and feature matching used for drone image stitching. In this paper, we compare the performance of three different feature extraction techniques such as SIFT (Scale-Invariant Feature Transform), SURF (Speeded-Up Robust Features), and ORB (ORiented FAST and rotated BRIEF) for detecting the key features. Then the detected features are matched using feature matching algorithms such as FLANN (Fast Library for Approximate Nearest Neighbors) and BF (Brute Force). All the matched key points may not be useful for creating panoramic image. Further, RANSAC (Random sample consensus) algorithm is applied to separate the inliers from the outlier set and interesting points are obtained to create a high-resolution image.

1 Introduction

Many technologies have been developed to produce high-resolution images with a wide view of the scene. However, they have limitation to capture the whole scene at an instance. Panoramic stitching is a technique used widely to overcome this problem. Image or panoramic stitching is the process of combining multiple overlapping images of different views to produce a high-resolution image. The process can be achieved in two ways: the direct pixel to pixel approach and feature-based approach [1]. The direct technique performs matches on each of the pixel to other pixel in

M. Dhana Lakshmi (✉) · P. Mirunalini · R. Priyadharsini · T. T. Mirnalinee
Department of Computer Science and Engineering, SSN College of Engineering, Chennai, India
e-mail: dhanalakshmi1605@cse.ssn.edu.in

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_59

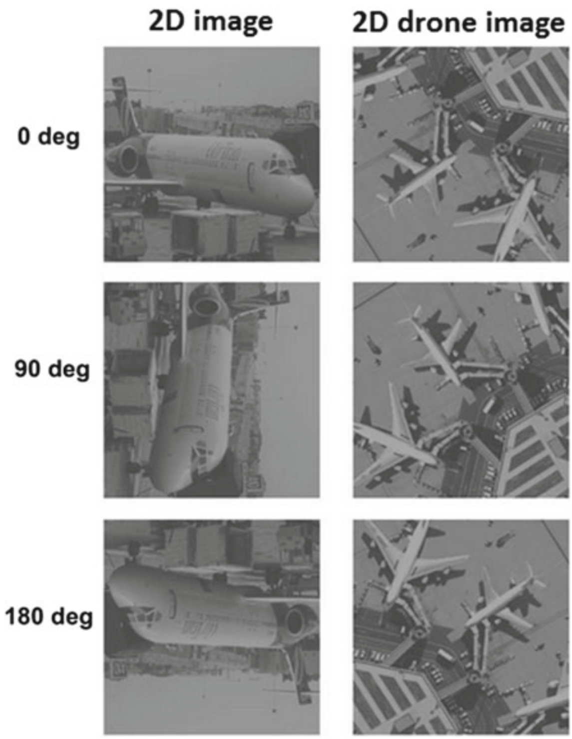


Fig. 1 2D images and 2D drone images with respect to different degree of rotation

order to reduce the mismatches. The feature-based technique, extract the features and perform the matches on the extracted features. Terrestrial scenes (2D images) captured by the digital handheld cameras and smart phones are significantly affected by rotation and become much more difficult to recognize the scene of an image. As the degree of rotation for a terrestrial scene increases, the harder it becomes to identify the scene information as shown in Fig. 1. In stitching of the terrestial scenes, the aerial perspective or top view cannot be retained or captured as such with respect to 2D images. Thus, stitching using 2D images will result in lack of information in top view. When 2D images are used to generate an image stitching of a building, then the roof parts and other structures that are visible from an aerial perspective will not be captured.

But image stitching can also be done using drone images captured by Unmanned Aerial Vehicle (UAV). Drone images contain the high aerial perspective, overlapping of an image and complete information of a scene of an image than the 2D terrestial images. These information helps the image stitching technique to build a high-quality image. The panoramic view of drone images obtained after stitching can be applied in the several fields such as movie industry, civil and mechanical industry.

2 Related Work

The image stitching has been implemented by many researchers in different ways. In [2] the authors reviewed the different feature-based image stitching methods such as SIFT, SURF, and ORB and also proposed a new method A-KAZE. It has been found among three methods that SURF-based visual odometry shows best accuracy for KITTI benchmark dataset. The proposed A-KAZE features demonstrated variation of motion estimation accuracy and computation efficiency. Parallel architecture for image fusion based on ORB feature identifier on a multicore DSP platform has been proposed in [3]. The methodology uses a position weighted image fusion algorithm to stitch the images. A panoramic image stitching technique for rotational images was proposed in [4], which used SIFT and SURF feature detector algorithms and blends the two images using DWT (Discrete Wavelet Transform) after obtaining the matches between the images. A comparison between different feature detector algorithm for the image stitching such as SIFT, SURF, ORB, FAST, Harris corner detector, FAST, MSER detector was done in [5].

3 Feature Extraction Techniques

Keypoints are the dominant features of an image. Features are contributed by the structures and the properties of an image such as color, texture, points, edges, objects, etc. Various feature extraction algorithm such SIFT, SURF and ORB were used to extract key features from drone images.

3.1 SIFT

The SIFT algorithm [6], is invariant to scaling and rotation of an image. It can also handle significant changes in illumination and efficient to run in real time. The SIFT algorithm can be achieved in four steps. It detects a scale space extrema, by generating the several octaves of the original image. Within an octave, images are progressively blurred using the Gaussian Blur operator. Two consecutive images in an octave are taken and one is subtracted from the other. Then the next consecutive pair is taken, and the process repeats. This is done for all octaves. The resultants are the approximation of Gaussian. Localization of keypoint is done by comparing neighboring pixels in the present scale, the successor scale and the predecessor scale. Keypoints can be rejected if they had a low contrast or if they were located on an edge. The assignment of orientation can be done by gradient directions and magnitudes around each keypoint. An image descriptor at each keypoint has been computed using a descriptor generator and stored as a descriptor [2, 7]. SIFT which is scale-invariant, extract the key features of image by resizing the image at different scales. So, all the important features have been extracted due to its invariant property. This property greatly enhances the degree of orientation and performs better in close range and aerial photography.

3.2 SURF

SURF algorithm is fast, robust feature detection and extraction algorithm. It approximates the Laplacian of Gaussian with Box filter and computes local extrema using second-order derivative. Implementation of Haar-like operators over an integral image can fasten the SURF in an efficient manner when compared to SIFT. For orientation assignment, it uses Haar wavelet by applying Gaussian weights for feature description. A keypoint may have connected neighbors which can be chosen and splitted into subparts. The wavelet responses are applied on each of the part to obtain feature descriptor. The features having same type of contrast can perform matching in faster rate [7, 8].

3.3 ORB

ORB is the combination of oriented FAST (Features from Accelerated Segment Test) and rotated BRIEF (Binary Robust Independent Elementary Fast) with some modification in order to enhance the performance of keypoint identification. FAST method is repeatedly applied to each layer of the pyramid in order to achieve the scale-invariant feature. The N keypoints which are computed based on the Harris corner measure are retained and uninteresting keypoints are eliminated. ORB adopts a rotation-aware variant of BRIEF. For detected keypoints, it finds patch centroids by image moments. The moments of a vector, links the keypoint's center to patch's centroid. The binary test pattern is rotated by the moments of a patch which allows feature to be in rotation-invariant form [8, 9].

4 Feature Matching Techniques

Keypoint matching is the process of finding correspondences between two images of the same scene or object. Drone images contain the 50–80% of overlapping between the images. Among the feature points extracted, the points that can be used for stitching are identified by the feature matching methods and the similar points from one image is mapped to points in the other image.

4.1 FLANN

FLANN is a library [10] of optimized algorithms that performs fast nearest neighbor search in high-dimensional features and large datasets. The FLANN uses randomized kd tree algorithm and does the priority search using k-means tree algorithm. Randomized kd-tree algorithm can search multiple trees in parallel by finding a point in the kd-tree which is nearest to a given input point [11]. The search can quickly eliminate the part of the search space by using the tree properties. Priority Search K-Means Tree Algorithm splits the data into M multiple regions and recursively partitioning each zone until the each of the leaf node has no more than M items. Then, picks up the initial centers in random manner [12].

4.2 Brute Force Matcher

BF matcher tries all possibilities and finds the best matches [10]. It takes the descriptor of a feature in an image and compared it using distance measure with all other features in the second image. The nearest point is represented as matched keypoints between the images. The steps involved in Brute force algorithm are as follows:

1. The distance between reference points (first image) and query points (second image) are calculated.
2. The calculated distances are sorted.
3. The *k*-smallest distances are selected as reference points.
4. The steps 1–3 is repeated for all query points.

FLANN considered to be faster since it compares between the nearest points of two images. But BF matcher compares every point in one image with all other point in another image.

5 Image Stitching Technique

An interest point (TP) is a specific location that is recognizable visually in the overlap area between two or more images. Interest points are also considered as inliers which is separated from the outliers to create an image matches.

RANSAC

RANSAC is used to separate matching keypoints (inliers) from non-matching keypoints (outliers) and create an image match. A part of data items is selected aimlessly from the input. An applicable model and the corresponding model parameters are computed using the items of this sample part. The algorithm checks the elements of the whole dataset whether it is consistent with the already represented model. A data element which does not fit the model is considered as outliers. The set of inliers obtained from the applicable model is called concord set. The RANSAC algorithm will run repeatedly until the obtained concord set in certain iteration has enough inliers (valid keypoints). This helps in generation of the interest points used to stitch the images. It also calculates the homography between the images in 3×3 matrix form. This homography matrix establishes the relationship between the two images in order to obtain the stitched image [1, 13–15].

6 Experimental Results

The sample overlapping drone images shown in Fig. 2 are taken from DJI building dataset. These images are captured by the UAV named AscTec Falcon 8 (Ascending Technologies) using Sony NEX-5 (RGB) camera.

The sample images consist of 50–80% of overlapping region. In this paper, feature-based image stitching have been done using Opencv. The methods used for the discussion are as follows:

i. Feature extraction—SIFT, SURF and ORB
ii. Feature matching—FLANN and BF
iii. Interest point generation—RANSAC.

A possible combination of feature extraction and feature matching methods with RANSAC has been performed and the results have been visually compared. The following combination of methods has been carried out and the results are shown below in Fig. 3.

(a) SIFT + FLANN + RANSAC
(b) SIFT + BRUTE FORCE + RANSAC
(c) SURF + FLANN + RANSAC
(d) SURF + BRUTE FORCE +RANSAC
(e) ORB + FLANN + RANSAC
(f) ORB + BF + RANSAC.

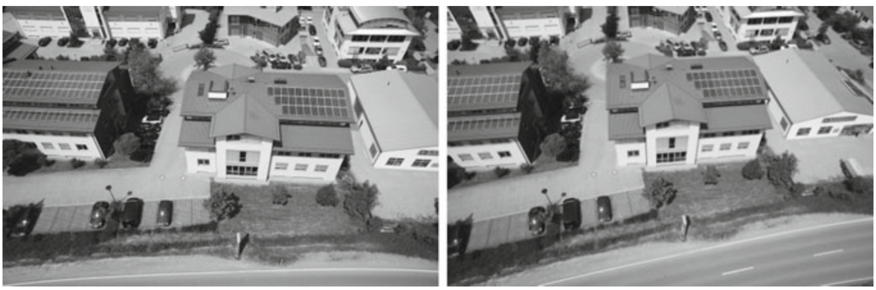


Fig. 2 Sample drone images

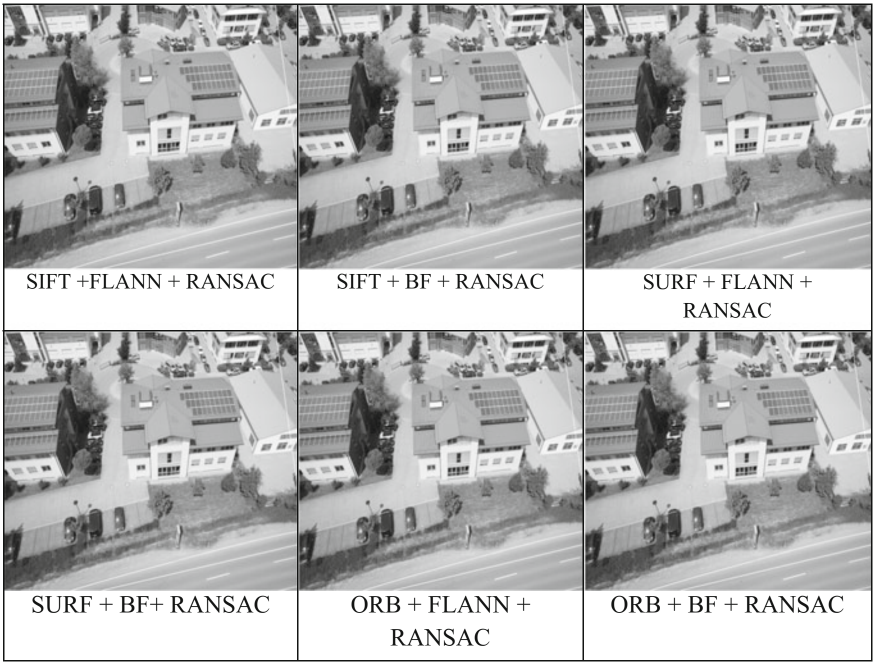


Fig. 3 Experimental results of various feature extraction and matching methods with RANSAC

7 Conclusions

In this paper, SIFT, SURF, and ORB algorithms are used for feature extraction as they efficiently detect the features in distinctive descriptor vector form. On the other hand, ORB extracts the features in binary string. Then feature matching algorithm FLANN and BF are performed on the extracted features. Though the features are matched between the two images of a scene, all matched features are not reliable that is they may not be an interest points used for stitching. RANSAC separates the valid inliers from outliers of matched features. It also calculates the homography between the images in 3×3 matrix form. This homography matrix establishes the relationship between the two images in order to obtain the stitched image. This work has justified that feature-based approach can be used for aerial drone images in image stitching which helps in full view of the scene.

References

1. Szeliski R (2006) Image alignment and stitching: a tutorial. Found Trends Comput Graph Vis 2(1):1–104
2. Chien HJ, Chuang CC, Chen CY, Klette R (2016) When to use what feature? SIFT, SURF, ORB, or A-KAZE features for monocular visual odometry. In: International conference on image and vision computing New Zealand
3. Wang G, Zhai Z, Xu B, Cheng Y (2017) A parallel method for aerial image stitching using ORB feature points. In: 2017 IEEE/ACIS 16th international conference on computer and information science
4. Bind VS, Muduli PR, Pati UC (2013) Robust technique for feature-based image mosaicing using image fusion
5. Adel E, Elmogy M, Elbakry M (2014) Real time image mosaicing system based on feature extraction techniques. In: 9th International conference computer engineering and systems (ICCES)
6. Lowe DG (2004) Distinctive image features from scale-invariant keypoints
7. Karami E, Prasad S, Shehata M (2017) Image matching using SIFT, SURF, BRIEF and ORB: performance comparison for distorted images
8. Rao T, Ikenaga T (2017) Quadrant segmentation and ring-like searching based FPGA implementation of ORB matching system for Full-HD video. In: 2017 Fifteenth IAPR international conference on machine vision applications, Nagoya University, Nagoya, Japan
9. Rublee E, Rabaud V, Konolige K, Bradski G (2011) ORB: an efficient alternative to SIFT or SURF. In: 2011 IEEE international conference on computer vision
10. Opencv.org. (2018) OpenCV | OpenCV. [online] Available at: http://opencv.org
11. Suju DA, Jose H (2017) FLANN: fast approximate nearest neighbour search algorithm for elucidating human-wildlife conflicts in forest areas. In: 4th International conference on signal processing, communications and networking
12. Ha Y-J, Kang H-D (2017) Evaluation of feature based image stitching algorithm using OpenCV. In: 10th International conference on (HSI) human system interactions
13. Steedly D (2005) Efficiently registering video into panoramic mosaics. Computer Science Department Faculty Publication Series 84
14. Tidke KS, Banarase SJ (2014) Review on image Mosaicing based on phase correlation and Harris Algorithm. Int J Adv Res Comput Sci Manage Stud 2(2)
15. Agarwal A, Jawahar CV, Narayanan PJ (2005) A survey of planar homography estimation techniques

An ANN-Based Detection of Obstructive Sleep Apnea from Simultaneous ECG and SpO_2 Recordings

Meghna Punjabi and Sapna Prabhu

Abstract Obstructive sleep apnea (OSA) is one of the most common sleep disorders characterized by a disruption of breathing during sleep. This disease, though common, goes undiagnosed in most cases because of the inconvenience, cost, and/or unavailability of opting for polysomnography (PSG) and a sleep analyst. Many researchers are working on devising an unsupervised, cost-effective, and convenient OSA detection methods which will aid the timely diagnosis of this sleep disorder. Commonly used signals to detect OSA are ECG, EEG, pulse oximetry (SpO_2), blood oxygen saturation (SaO_2), and heart rate variability (HRV). In this work, an attempt to detect the OSA using simultaneously acquired ECG and SpO_2 signals has been presented. Various features from the RR intervals of ECG, and a couple of features—-namely, CT90 and delta index—from the SpO_2, were extracted as indicators of OSA. The features were then fed to a trained artificial neural network (ANN) which classified the signals as OSA positive or OSA negative. The proposed technique boasts a very high accuracy of 98.3%, which is superior to other competing techniques reported so far.

1 Introduction

A balanced sleep is essential for normal functioning of the human body. Sleep disorders disrupt the ability to sleep well and can potentially cause physical, emotional, and psychological damage. Among the numerous sleeping disorders currently known, the most common ones are insomnia, narcolepsy, sleep apnea, and restless leg syndrome. Sleep apnea is characterized by disruption in a person's breathing during sleep, breathing might get shallow, or even stop for few seconds. The breaks during

M. Punjabi (✉) · S. Prabhu
Department of Electronics Engineering, Fr. Conceicao Rodrigues College of Engineering, University of Mumbai, Bandra 400050, Mumbai, India
e-mail: meghna_1911@hotmail.com

S. Prabhu
e-mail: sapna@frcrce.ac.in

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_60

breathing called apnea could be of different frequency and duration. If left untreated could cause repeated interruption while sleeping. There are three types of sleep apnea: obstructive (OSA), central (CSA), and a combination of the two called mixed. The most common form is the first type OSA found in 2–4% of middle-aged adults and 1–3% of children. OSA is caused by a collapse of the upper respiratory airway. The CSA occurs due to inhibited or absent respiratory drive. In general, the mixed type is found, and the occurrence of CSA only cases is quite rare. Although they are commonly present, diagnosing can be challenging. The reason behind undiagnosed cases in sleep apnea is due to the inconvenience, cost, and unavailability of testing. Current diagnosis for testing sleep apnea is done using polysomnography (PSG). PSG is a standard test for all sleeping disorders, the breath airflow, respiratory movements, oxygen saturation, body position, electroencephalogram (EEG), electrooculogram (EOG), electromyogram (EMG), and electrocardiogram (ECG).

There are different methods proposed for diagnosing sleep apnea till date. Widely used statistical parameters for detecting sleep apnea are nasal airflow, thorax and abdomen effort signals, acoustic speech signal, oxygen saturation, electrical activity of the brain (EEG), and electrical activity of the heart (ECG).

The detection technique for sleep apnea using thoracic and abdominal signals was reported by Ng et al. [13]. They used mean absolute amplitude analysis and showed that the combination of the thoracic and abdominal signal had the best overall and individual performance when compared to the separate performances of the thoracic and abdominal signal.

Goldshtein et al. [9] hypothesized that the patients with OSA would exhibit different speech signal properties than those without OSA. They had studied acoustic speech features of 93 subjects recorded using text-dependent speech protocol and a digital audio recorder before doing polysomnography study. With the help of a Gaussian mixture model-based system, they had developed a model and classified based on the features like vocal tract length and linear prediction coefficients. It was concluded that the acoustic features from speech signals during wakefulness can detect OSA patients with good specificity and sensitivity.

Another method of distinguishing OSA positive and negative patients was done using nocturnal oximetry which was superior to polysomnography with respect to its low cost and simplicity. The nocturnal oximetry assesses nonlinear analysis of blood oxygen saturation (SaO_2). The oximetric indices evaluated were cumulative time spent, oxygen desaturation, and delta index. The result of the nonlinear analysis of SaO_2 signals [4] from nocturnal oximetry suggested that it could lead to useful information in OSA diagnosis.

Another method of identifying sleep apnea episodes was reported by Lin et al. [11]. This method uses artificial neural networks to extract EEG signal characteristics of sleep apnea episodes. They used wavelet neural network as a sleep apnea identification system. They were able to achieve a sensitivity of approximately 69.6% and specificity of 44.4%.

There have been several methods for detection of sleep apnea through heart rate variability analysis. For example, Manrique et al. [16] used time–frequency distribution to extract dynamic features for detecting sleep apnea that are recorded from

ECG signals during sleep. They proposed a method that can be used as a simple diagnostic tool for sleep apnea with an accuracy of 92.7% in one-minute intervals. Another work on detecting OSA using ECG [17] in which analysis and annotations are based on spectral components of heart rate variability, frequency analysis performed using Fourier and wavelet transformation with appropriate application of Hiber transform were able to achieve a sensitivity of 90.8% and a specificity of 92.7% on the learning set. Mendez et al. [12] showed that a bivariate autoregressive model used to evaluate beat-by-beat power spectral density of HRV and R peak area had higher than 85% accuracy in the classification result. The model is based on extraction of signal ECG signal characteristics. An improved technique based on automated classification algorithm was explored in Almazaydeh et al. [1]. This technique processed short duration epochs of ECG data. The automated classification algorithm was based on support vector machines (SVM). The resultant automated classification system developed showed a high degree of accuracy approximately 96.5% in recognizing epochs of sleep disorder.

Xie et al. [18] explored various features in order to find an efficient alternative for polysomnography (PSG). They investigated real-time sleep apnea and hypopnea syndrome (SAHS) based on ECG and SpO_2 signals both separately and combined. They showed that the SpO_2 with the proposed features outperformed ECG in terms of diagnostic capability.

In this work, we are proposing a novel ANN-based technique for the detection of OSA using the features derived from simultaneously acquired ECG and SpO_2 signals, which is an improvement on the previously reported techniques.

2 Methods

An attempt to detect OSA using simultaneous ECG and SpO_2 signals acquired from the patients during the sleeping phase is presented in this work. In short, the acquired signals were divided into segments of 1 min duration. For each of those segments, the features obtained from the ECG and SpO_2 were combined to form a feature vector, which was given as the input to an ANN. The ANN is trained to classify the feature vector into OSA positive or OSA negative. Using the annotated data obtained from the PhysioNet Databank [15], the performance of the proposed OSA detection technique was evaluated. Each of the steps involved in the method is described in the subsections below.

2.1 Preparation of the Dataset

The "Apnea-ECG" database [15] of "PhysioNet" [8] contains recordings from 70 patients—which include a continuous digitized ECG signal, a set of apnea annotations derived by human experts on the basis of simultaneously recorded respiration

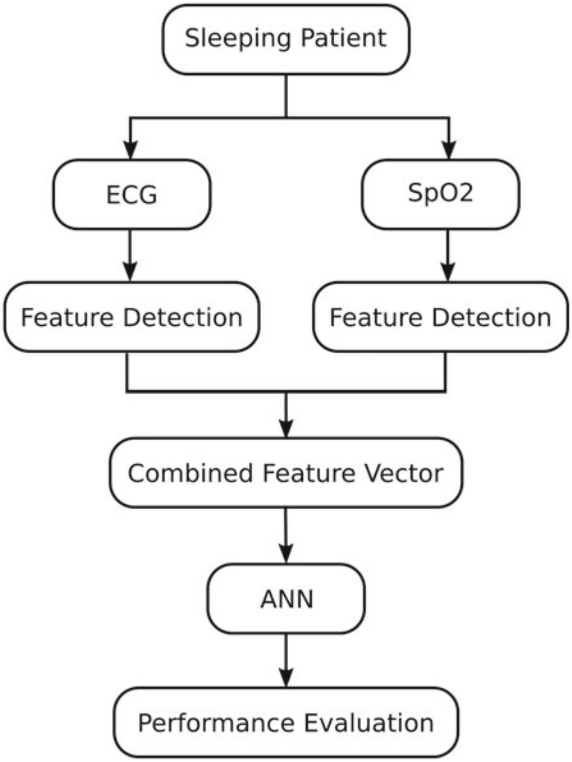


Fig. 1 Schematic diagram of the OSA detection system

and related signals, and a set of machine-generated QRS annotations in which all beats regardless of type have been labeled normal. Among those 70 recordings, eight recordings (a01 through a04, b01, and c01 through c03) are accompanied by four additional signals (Resp C and Resp A, chest and abdominal respiratory effort signals obtained using inductance plethysmography; RespN, oronasal airflow measured using nasal thermistors; and SpO_2, oxygen saturation). As our technique makes use of simultaneous ECG and SpO_2 recordings, we restrict our study to these eight recordings.

The ECG and SpO_2 records are present in recording Channels 1 and 5, respectively. Annotations done by expert physicians are available for each of the shortlisted records. These annotations are done at the beginning of every minute of the recording. As the sleeping sessions are roughly 8 h long, there is around 500 min of recordings available. Every minute of the recordings was treated as an individual signal segment (Fig. 1).

The annotations which are present at the beginning and end of the segments were used to label the segments as OSA positive and OSA negative. The following rules were used:

1. If the segment is bounded by two "N" annotations, it represents a normal segment or N-Seg.
2. If the segment is bounded by two "A" annotations, it represents an apnea segment or A-Seg.
3. If the segment is bounded by one "N" annotation and one "A" annotation, the status of the patient is ambiguous during this segment and hence these segments were not used in the study.

Once the segments were identified, the key features of the segments which aid the detection of OSA were identified. The features used in this study are described in the following subsection.

2.2 Feature Detection

For every segment of sleep, there is a corresponding ECG and SpO_2 segment. The features of ECG and SpO_2 were separately evaluated.

The ECG features used were all based on the variations in RR interval observed during OSA [1, 5, 19]. As the R peaks of the ECG signals from Apnea-ECG database were not annotated, those were detected using the well-established Pan-Tompkins algorithm [14]. Once the R peak locations were identified, the RR intervals for the ECG segment were obtained. The features extracted from the series of RR intervals corresponding to the segment of interest are given below:

1. Mean RR interval,
2. Standard deviation (SD) of RR interval,
3. NN50 variant 1:The number of RR interval pairs in which the first interval exceeds the second by at least 50 ms,
4. NN50 variant 2: The number of RR interval pairs in which the second interval exceeds the first by at least 50 ms,
5. pNN50 variant 1: NN50v1 represented as a fraction of total number of RR intervals in the segment,
6. pNN50 variant 2: NN50v2 represented as a fraction of total number of RR intervals in the segment,
7. SDSD: Defined as "the standard deviation of the differences between the adjacent RR intervals" [1],
8. RMSSD: Defined as "the square root of the mean of the sum of the squares of differences between adjacent RR—intervals" [1],
9. Median of the RR intervals,
10. Interquartile range of the RR intervals,
11. MADV: Mean absolute deviation values, defined as "the mean of absolute values obtained by the subtraction of the mean RR-interval values from all the RR-interval values in an epoch" [1], and

12. Spectral Ratio: defined as the ratio of the area under the power spectrum of the RR interval between the frequency bands (0.02–0.1) and (0.01–0.02), relative to the Nyquest frequency [6].

The features from the SpO_2 segments suitable for OSA detection were defined in the earlier works [2, 3]. Two of those features, namely CT90 and delta index, were used for the work presented here. These features were defined as follows:

1. CT90: The cumulative time spent below 90% oxygen saturation level is quantified by the feature CT90. In [3], CT90 is defined as the percentage of time during which the SpO_2 value is below 90%.
2. Delta Index: The factor delta index quantifies the SpO_2 variability. This feature was originally defined in [10] and also used by recent research works such as [2] and [7]. The metric delta index is defined as the average of absolute differences of the mean SpO_2 between successive 12-s intervals [7].

2.3 Preparation of the Feature Vector

Once these features values were evaluated from the ECG and SpO_2 segments, they were combined to form a feature vector. The feature vector is a series of 14 values, each of which represents individual features—12 ECG features and 2 SpO_2 features—in order. It is assumed that this feature vector contains all the relevant information from the raw ECG and SpO_2 segments required for the detection of the OSA.

2.4 Detection of OSA Using ANN

The individual features extracted from the ECG and SpO_2 segments vary in their values for normal and apneatic ECG/SpO_2 records. The extent to which each of the features expresses their variation sensitive to the presence or absence of OSA is different for different features. A machine learning algorithm was designed to classify the sleep segment as Normal or Apnea using a neural network classifier which learns the appropriate weights of each of the features by making use of a training set.

For the application presented in this work, the minimum number of neurons needed in the input is the length of the feature vector, i.e., 14. The number of output layer neurons is also fixed, which is unity, as a Boolean output (0 or 1) is required which can be provided by a single neuron. Our choice in deciding the network topology lies in the hidden layer, where the number of layers and the number of neurons in each layer can be decided by the user. We have chosen for one hidden layer with 10 neurons. The Neural Fit Tool (NFTOOL) of the NN toolbox available in the Matlab platform was used to create the neural network design.

The individual neuron in the network was assigned a log-sigmoid (logsig) transfer function. The inputs to the neuron are scaled using a weight assigned to the neuron. The process of optimizing these weights so that the network produces desirable outputs for all input vectors is the training process. The Levenberg–Marquardt algorithm (LMA) was used as the training function, which computes the optimized neuron weights which produce the least mean square error (LMS error) between the NN outputs and the TARGETS, where TARGETS are the desired output for each vector in the training set.

Once the training is completed, the network can be used to detect the sleep apnea from a pair of ECG and SpO_2 segments by providing the 14-element feature vector as the input. The logsig transfer function of the output neuron provides a real number between 0 and 1. Hence a rounding off operation is carried out on the ANN output for the ease of apnea detection, which makes the output Boolean, representing OSA positive and OSA negative detection.

2.5 Performance Evaluation

Multiple test sets were prepared separately from different patients. Individual samples in these datasets were taken through the OSA detection procedure explained in the previous section. From the ground truth annotations available from the database and the outputs obtained from the trained ANN, the performance indicators such as sensitivity, specificity, and accuracy were measured. The confusion matrix which displays detailed performance indicators were also obtained for each of the datasets. The results thus obtained were then compared with the results from the similar work found in the literature.

3 Results

Eight datasets, one each from individual patients, were prepared for the training of the ANN. Each dataset consists of multiple ECG—SpO_2 segment pairs of 1 min duration. Each segment pair was considered as one training sample. The number of samples was limited to 200 from each patient, which were randomly chosen from the entire recording. The details of the datasets used are given in Table 1. A combined dataset containing the pool of all patient-specific datasets was used to train the ANN.

The samples present in the above data pool were randomly grouped into three categories in the following manner:

1. Training Set: 70% of the entire samples are chosen to represent the Training Set. This was the set used to train the ANN parameters. For each segment-pair in the Training Set, a feature vector was constructed as mentioned in the methods. These feature vectors were given as the INPUTS to the ANN along with the

Table 1 Details of the data sets which were used to train the ANN in detecting the presence of OSA

Dataset	Patient	Total segments	OSA +ve	OSA −ve
Set-1	a01er	150	135	15
Set-2	a02er	150	75	75
Set-3	a03er	200	100	100
Set-4	a04er	200	165	35
Set-5	b01er	63	13	50
Set-6	c01er	200	0	200
Set-7	c02er	200	0	200
Set-8	c02er	200	0	200

corresponding annotations as TARGETS which indicate the desired output of the network.

2. Validation Set: 15% of the samples were chosen to represent the Validation Set, which is used for cross-validation. There is a possibility that the NN overfit the training set and generate 100% accuracy for that training set. A validation set was used for avoiding this overfit. If the ANN gets overfit to the training data, it gives a very low accuracy for the validation dataset. In such cases, the training process was repeated with a different set of values for the NN weights. This process was repeated until a similar performance was obtained for training dataset and the validation dataset.
3. Test Set: The final 15% of the samples were used as test data. After the entire training process, these data will be used to evaluate the efficiency of the apnea detection.

The training process was repeated until a maximum accuracy of 94.5% was obtained for the Test Set. In order to compare the performance of our technique to previously published techniques, a test dataset used in the work of Almazaydeh et al. [1] was chosen for detailed evaluation of the proposed technique. Almazaydeh dataset was taken from recording a03er. For regular (OSA −ve) samples, the recording stretch. The confusion matrix obtained by the proposed method. A test data set containing 58 samples was used. TP, TN, FP, and FN represent True Positives, True Negatives, False Positives, and False Negatives, respectively. From 2:27:00.000 to 2:57:00.000 was used. For apnea (OSA +ve) samples, the stretch from 3:06:00.000 to 3:36:00:000 was used. As our samples are of 1 min duration, we have total 58 samples—29 each for OSA +ve and OSA –ve samples. For this dataset, the proposed technique was able to yield 98.3% accuracy with 100% specificity and 96.6% sensitivity (Fig. 2) which is superior to the accuracy achieved by any other technique reported in the literature so far. Comparison of the proposed technique with similar prominent works published earlier are shown in Table 2.

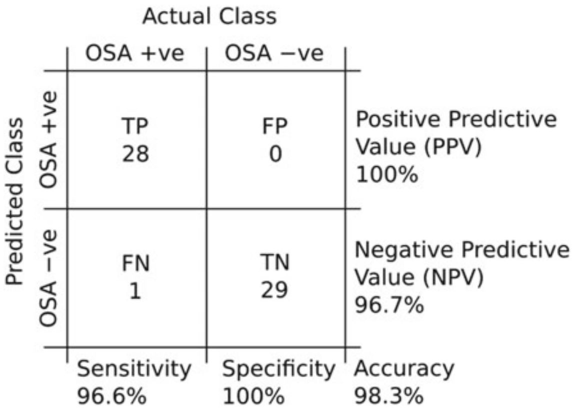


Fig. 2 The confusion matrix obtained by the proposed method. A test data set containing 58 samples was used. TP, TN, FP, and FN represent True Positives, True Negatives, False Positives, and False Negatives, respectively

Table 2 Comparison of performances in OSA detection approaches

Method by	Refs.	Signal(s) used	Performance (%)		
			Se	Sp	Acc.
Schrader et al.	[17]	Fourier and wavelet transformation of HRV	90.8	NA	NA
Chazal et al.	[5]	Measure of minutes of sleep disordered respiration	NA	NA	91
Lin et al.	[11]	EEG	69.6	44.4	NA
Alvarez et al.	[3]	SaO_2	90.1	82.9	NA
Mendez et al.	[12]	Bivariate autoregressive model of HRV	NA	NA	85
Alvarez et al.	[4]	SaO_2 and EEG	91	83.3	88.5
Manrique et al.	[16]	ECG	NA	NA	92.7
Yilmaz et al.	[19]	RR interval-based classification	NA	NA	89
Xie et al.	[18]	SpO_2 and ECG	79.7	85.9	84.4
Almazaydeh et al.	[1]	ECG	92.9	100	96.5
Proposed		SpO_2 and ECG	96.6	100	98.3

4 Conclusions

This work proposes an ovel method for the detection of OSA using the features detected from simultaneously acquired ECG and SpO_2 recordings from apnea patients during sleep. A machine learning algorithm based on ANN was designed and trained to detect OSA using the shortlisted features. The performance analysis of the technique shows that it is superior to similar algorithms designed so far for unsupervised OSA detection.

The two signals used in the proposed techniques are ECG and SpO_2—two non-invasive, cheap, and commonly available instruments in hospitals and affordable in home. The patients can easily acquire these signals and, with the help of the proposed technique, can get the diagnosis done for OSA. In future, this technique can be incorporated into a real-time ECG + SpO_2 acquisition system and aid the process of sleep analysis.

Acknowledgements We thank our institute, Fr. Conceicao Rodrigues College of Engineering, for providing all possible supports to this work.

References

1. Almazaydeh L, Elleithy K, Faezipour M (2012) Obstructive sleep apnea detection using SVM-based classification of ECG signal features. In: 2012 Annual international conference of the IEEE engineering in medicine and biology society (EMBC). IEEE, pp 4938–4941
2. Almazaydeh L, Faezipour M, Elleithy K (2012) A neural network system for detection of obstructive sleep apnea through SpO_2 signal. Editorial Preface 3(5)
3. Alvarez D, Hornero R, Abasolo D, Del Campo F, Zamarron C (2006) Nonlinear characteristics of blood oxygen saturation from nocturnal oximetry for obstructive sleep apnoea detection. Physiol Meas 27(4):399
4. Alvarez D, Hornero R, Marcos JV, del Campo F, Lopez M (2009) Spectral analysis of electroencephalogram and oximetric signals in obstructive sleep apnea diagnosis. In: Annual international conference of the IEEE engineering in medicine and biology society. EMBC 2009. IEEE, pp 400–403
5. de Chazal P, Penzel T, Heneghan C (2004) Automated detection of obstructive sleep apnoea at different time scales using the electrocardiogram. Physiol Meas 25(4):967
6. Drinnan M, Allen J, Langley P, Murray A (2000) Detection of sleep apnoea from frequency analysis of heart rate variability. In: Computers in Cardiology 2000. IEEE, pp 259–262
7. Garde A, Dehkordi P, Wensley D, Ansermino JM, Dumonf GA (2015) Using oximetry dynamics to screen for sleep disordered breathing at varying thresholds of severity. In: 2015 23rd European signal processing conference (EUSIPCO). IEEE, pp 439–443
8. Goldberger AL, Amaral LA, Glass L, Hausdorff JM, Ivanov PC, Mark RG, Mietus JE, Moody GB, Peng CK, Stanley HE (2000) Physiobank, physiotoolkit, and physionet. Circulation 101(23):e215–e220
9. Goldshtein E, Tarasiuk A, Zigel Y (2011) Automatic detection of obstructive sleep apnea using speech signals. IEEE Trans Biomed Eng 58(5):1373–1382
10. Levy P, Pepin JL, Deschaux-Blanc C, Paramelle B, Brambilla C (1996) Accuracy of oximetry for detection of respiratory disturbances in sleep apnea syndrome. Chest 109(2):395–399

11. Lin R, Lee RG, Tseng CL, Zhou HK, Chao CF, Jiang JA (2006) A new approach for identifying sleep apnea syndrome using wavelet transform and neural networks. Biomed Eng Appl Basis Commun 18(03):138–143
12. Mendez MO, Ruini DD, Villantieri OP, Matteucci M, Penzel T, Cerutti S, Bianchi AM (2007) Detection of sleep apnea from surface ECG based on features extracted by an autoregressive model. In: 29th Annual international conference of the IEEE engineering in medicine and biology society. EMBS 2007. IEEE, pp 6105–6108
13. Ng AS, Chung JW, Gohel MD, Yu WW, Fan KL, Wong TK (2008) Evaluation of the performance of using mean absolute amplitude analysis of thoracic and abdominal signals for immediate indication of sleep apnoea events. J Clin Nurs 17(17):2360–2366
14. Pan J, Tompkins WJ (1985) A real-time QRS detection algorithm. IEEE Trans Biomed Eng 32(3):230–236
15. Penzel T, Moody GB, Mark RG, Goldberger AL, Peter JH (2000) The apnea-ECG database. In: Computers in cardiology 2000. IEEE, pp 255–258 (2000)
16. Quiceno-Manrique A, Alonso-Hernandez J, Travieso-Gonzalez C, Ferrer-Ballester M, Castellanos-Dominguez G (2009) Detection of obstructive sleep apnea in ECG recordings using time-frequency distributions and dynamic features. In: Annual international conference of the IEEE engineering in medicine and biology society. EMBC 2009. IEEE, pp 5559–5562
17. Schrader M, Zywietz C, Von Einem V, Widiger B, Joseph G (2000) Detection of sleep apnea in single channel ECGs from the PhysioNet data base. In: Computers in Cardiology 2000. IEEE, pp 263–266 (2000)
18. Xie B, Minn H (2012) Real-time sleep apnea detection by classifier combination. IEEE Trans Inf Technol Biomed 16(3):469–477
19. Yilmaz B, Asyali MH, Arikan E, Yetkin S, Ozgen F (2010) Sleep stage and obstructive apneaic epoch classification using single-lead ECG. Biomed Eng Online 9(1):39

Design of an Image Skeletonization Based Algorithm for Overcrowd Detection in Smart Building

R. Manjusha and Latha Parameswaran

Abstract Crowd analysis has found its significance in varied applications from security purposes to commercial use. This proposed algorithm aims at contour extraction from skeleton of the foreground image for identifying and counting people and for providing crowd alert in the given scene. The proposed algorithm is also compared with other conventional algorithms like HoG with SVM classifier, Haar cascade and Morphological Operator. Experimental results show that the proposed method aids better crowd analysis than the other three algorithms on varied datasets with varied illumination and varied concentration of people.

1 Introduction

Video analytics has gained its prevalence over recent years due to its flexibility and reduction in cost for the overall system. People counting and crowd analysis are active areas of research with the rapid increase in surveillance videos. Crowd alert is inevitable in public scenarios which demands crowd-induced disasters. Several disasters like the recent stampede at Mumbai's Elphinstone Road station on September 29, 2017, in which 22 were dead and several were injured, happens due to lack of proper crowd analysis and crowd alert. Crowd alert in smart building aids in providing emergency evacuation, analyzing abnormal crowd at unusual time, etc.

This paper focuses on counting the number of people to estimate and provide an alert if it is overcrowded, using skeleton-based contour extraction and compare it with three algorithms like HOG with SVM, Haar Cascade, and Morphological operator based algorithm. The paper is organized as follows: Sect. 2 discusses the existing algorithms for people counting and crowd analysis; Sect. 3 elaborates on the

R. Manjusha (✉) · L. Parameswaran
Department of Computer Science and Engineering, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Coimbatore, India
e-mail: r_manjusha@cb.amrita.edu

L. Parameswaran
e-mail: p_latha@cb.amrita.edu

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_61

proposed method and the three conventional methods with which the proposed work is compared; Sect. 4 provides the comparison and analysis of the proposed method with the other three algorithms.

2 Related Works

Many authors have published work on people counting. A brief literature is presented here.

In [1], Scale-Invariant Feature Transform (SIFT) was used for blob analysis. Complex wavelet transform (CWT) and dimensionality reduced Optical flow was applied to the detected blobs for estimating the movement direction and speed of the key points. If the result of CWT and Optical flow contains the blob, it is treated as a person. The people map generated was then applied with Kalman filter for tracking and thus occlusion was dealt. In [2] authors used head shoulder region for identification of people; Histogram of Gradients (HoG) combined with Completed Local Binary Pattern (CLBP) features which were provided to Support Vector Machine (SVM) for identifying people.

A hybrid face detector [3] a combination of Normalized Pixel Difference (NPD), Haar classifier, and Haar classifier for profile face was used. If all the three detectors, NDP which detects faces at different scales and sizes, Haar cascade and Haar cascade for profile faces which works well for profiled and rotated faces identifies a region as a face then it is considered to be a face. False positives are further eliminated based on the area of the bounding box. In [4] the camera was placed overhead and people counting has been done based on the headcount. ViBe algorithm has been used for foreground extraction followed by closing and opening for removal of noise and holes in the extracted foreground. Local Binary Pattern (LBP) based Adaboost classifier has been used for detecting human heads. Head tracking has been further done based on mean shift algorithm. In [5] authors used Mixed Gaussian background model (MGM) for extraction of foreground, followed by dilation and erosion for removal of noise and blur. Histograms of Gradient (HoG) features were extracted and Principal Component Analysis (PCA) has been applied on HoG features to reduce the dimension of the feature vector. Support Vector Regression (SVR) has been used for training and identification of people.

In [6] authors have developed a Frame differential method along with dilation and erosion has been used for extracting foreground of each frame. Each connected component in the resulting foreground was treated as a group. Each frame has been divided into blocks and the area, the number of edge pixels and number of SURF points of each group falling in each block has been estimated. Support Vector Regression algorithm has been used for training and for predicting the number of people in each frame. A contracted graph with an entry point of people as a source (S) and the exit point of people as a sink (T) for a sequence of frames in a video has been constructed. The prediction of the regression algorithm has been improvised by using an integer quadratic programming model on the contracted graph.

Lidar has been placed vertically in [7] for collecting the 2D points which indicated the contours of each frame. These 2D points obtained from various frames were superimposed to identify contour over time. RANdom SAmple Consensus (RANSAC) algorithm has been further used for registration of these images which in turn creates a point cloud. Since the pedestrians too close to each other may be grouped into one cloud, the head shoulder information has been used for isolating people. SVM has been further used for detection and classification of people. Perng et al. [8] aim at counting the number of people in the bus at a point of time, based on the entry and exit count at each stop. Background subtraction intending to identify the moving objects as foreground has been done based on frame differencing. Erosion and dilation have been done to remove noise and broken edges. Connected component analysis based on area, aspect ratio, and velocity was used as a means to identify people. A virtual counting line and a counting zone have been set for tracking and counting the number of people who enter and exit the bus.

In [9] the authors focus on counting the number of people from ATM surveillance video. Static background has been subtracted from each image frame and based on a threshold moving target has been identified. Gaussian Model has been further used for extracting foreground pixels from the thresholded image. The resulting image was eroded and then dilated using a 2×7 structuring element. A rectangle was drawn for each contour and to identify people close to each other, the height and width of the rectangle have been considered. SURF (Speeded-Up Robust Features) has been used by [10] for identifying the interest points on blob-like structures throughout the image. From the set of interest points, those which corresponds to moving objects indicates a person. Adaptive Rood Pattern Search (ARPS) algorithm has been used for identifying those motion vectors. Training and recognition of people have been done using linear regression. In [11] an algorithm using head-based people counting has been implemented using Statistically Effective Multi-scale Block Local Binary Pattern (SEMB-LBP) features and Adaboost classifier. The region of interest which was 0.7–1.3 times larger than the human head has been initialized for counting the heads of people. A modified version of compressive tracking (CT) tracker based on model matching was used for tracking the detected head in the region of interest. In [12] saliency map had been generated based on the color and luminance information of the image. Object segmentation had been done based on dynamic mode decomposition. In [13] based on facial features like eyes, nose, and mouth of the person and using K-Nearest Neighbor classifier people identification had been done. A summary of literature helps to infer that people counting and overcrowding identification is still an open research.

3 Proposed Work

The proposed algorithm counts the number of people and with a user-defined threshold identifies overcrowd, if any, in each frame of a video using skeleton-based contour extraction. The proposed algorithm is compared with the standard existing algorithms

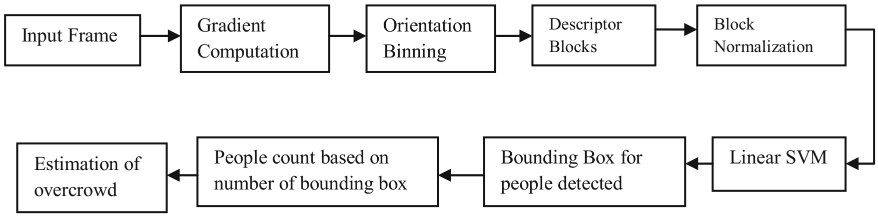


Fig. 1 Architecture of HOG with SVM

like HoG with SVM, Haar cascade, and morphological operator based algorithm. Existing algorithms and by the proposed algorithm for people counting overcrowding identification is presented here.

3.1 HOG with SVM

Histogram of oriented gradients (HOG) descriptor, divides the image (each frame in a video) into cells and histogram of gradient directions are computed for every pixel in each cell. The local histogram intensities of cells in a block are further normalized. Figure 1 shows the architecture of HOG descriptor with SVM algorithm [14]:

Algorithm:

a. Gradient Computation:

The first step involved is the computation of Gradients for each pixel in the frame. Gradients are computed using the standard derivative mask [−1 0 1], $[-1\ 0\ 1]^T$ in one direction or both horizontal and vertical direction.

b. Orientation Binning:

The next step is to calculate the cell histograms. Histogram contains channels ranging from 0° to 180° which corresponds to unsigned gradients. Depending on the gradient computed for each pixel, the corresponding histogram channels vote is updated.

c. Descriptor Blocks:

Cells are further grouped into larger blocks that are spatially connected which are then normalized. The normalized cell histograms are concatenated to form the HOG descriptor for each block.

d. Block Normalization:

The blocks are normalized using L1 or L2 normalization. This Normalized HOG descriptor is provided to SVM classifier for differentiating humans from non-humans.

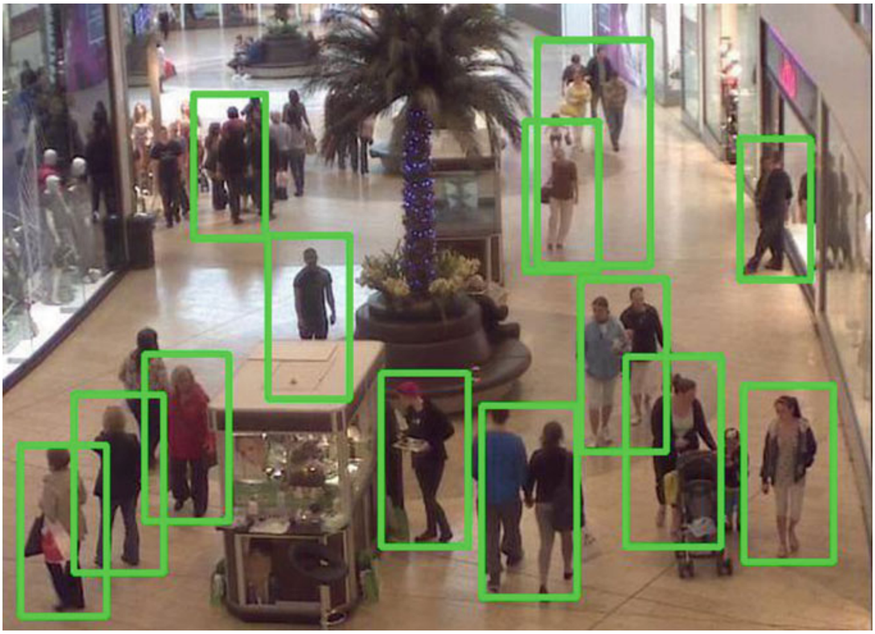


Fig. 2 HOG with SVM on Mall dataset

e. SVM Classifier:

HOG Descriptors is further used for detecting human which aids in people counting by providing those features to SVM (Support Vector Machine). HOG features are extracted both from positive and negative samples in the dataset. SVM is trained with all the samples and from each input video frame, the humans are detected based on the HOG descriptor of that frame and the trained samples. A rectangular bounding box is drawn around each detected human and is further used for counting the number of people in each frame. Based on user-defined threshold overcrowd is alerted.

Figures 2 and 3 shows the result of HOG with SVM on Mall dataset and Dataset from a surveillance camera of our University.

3.2 *Haar Cascade*

Haar feature-based cascade classifiers [15] represent each frame of a video as an intermediate representation known as an integral image. Integral image is of the same size as the original image. The sum of all pixels on the up-left region of the original image constitutes each element in an integral image. Sum = $I(C) + I(A) - I(B) - I(D)$ for a particular position as shown in Fig. 4.

Fig. 3 HOG with SVM on dataset from surveillance camera of our university

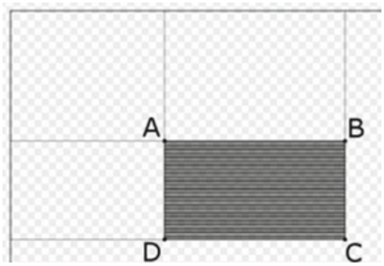


Fig. 4 Integral image

Fig. 5 Best features for face detection

From the integral image the features shown in Fig. 5 are extracted which suits best for detecting faces. The first feature is based on the fact that region of eyes is mostly darker than the cheeks and the nose region. The second feature is supported by the fact that eyes are darker than the nose bridge.

These features are extracted from the training images and an Adaboost classifier which is a combination of several weak classifiers classifies the face and the non-face

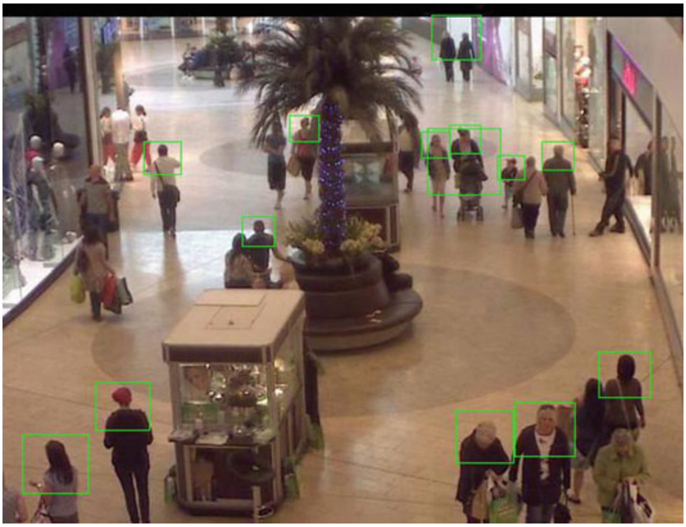


Fig. 6 Haar cascade on Mall dataset

Fig. 7 Haar cascade on dataset from surveillance camera of our campus

images. A cascade of classifiers has been used where only if the first classifier returns a positive result, the image shall be passed on to the second classifier, otherwise discarded and so on, thus minimizing false negatives.

Using the pre-trained Haar classifier for upper body, people are detected, which aids in people counting and overcrowd detection. Figure 6 shows the people detected from the Mall dataset and Fig. 7 shows the people detection on the dataset from the surveillance camera of our university.

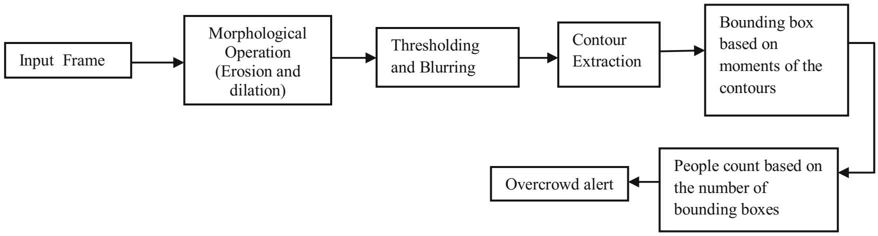


Fig. 8 Architecture of morphological operator based people counting

Fig. 9 Morphological operator based people counting on Mall dataset

3.3 Morphological Operators

Gaussian Mixture Model is used to separate the background from each frame of the video. On the foreground image of each frame, a series of erosion followed by a series of dilation are applied. Resultant images are threshold followed by median blurring. Contours are extracted from the blurred image by compressing the horizontal, vertical, and diagonal segments and leaving behind only the endpoints based on chain codes and only the outer contours are extracted. Based on the moments extracted from the contours, the rectangular bounding box is drawn on each region identified to be the person in the image. Crowd alert is provided based on the user-defined threshold. Figure 8 shows the architecture of morphological operators based people counting.

Figure 9 shows the result of morphological operator based people counting on Mall dataset and Fig. 10 shows the result of morphological operator based people counting on surveillance dataset of our university.

Fig. 10 Morphological-based people counting on dataset from surveillance camera of our campus

3.4 Proposed Skeletonization Based People Counting

Each frame of the video is applied with a Mixture of Gaussian Model for background separation. The extracted foreground image is converted to binary and then skeletonization is done on the binary image. An opening morphological operator which is erosion followed by dilation, where, erosion aids in removing smaller elements and dilation helps in restoring the shape of the remaining skeleton is then applied on the skeleton. The resultant skeleton is further smoothened and edges are detected using a canny edge detector. From the resulting skeleton, contours are extracted based on the chain codes by compressing the horizontal, vertical, and diagonal segments and leaving behind only the endpoints. Moments are calculated for the extracted contours and a bounding box is drawn on each person detected. The number of bounding box corresponds to the number of people in each frame. Based on the user-defined threshold, overcrowding is estimated. Figure 11 shows the architecture of people counting based on skeletonization.

Figure 12a shows the skeleton of the 1997th frame of the Mall dataset. Figure 12b shows the Skeletonization-based people counting algorithm applied on the video and the result on the 1997th frame of the Mall dataset.

Figure 13a shows the skeleton of 223rd frame and Fig. 13b shows the result of Skeletonization-based people counting algorithm applied on the same frame of the video from the surveillance camera of our university.

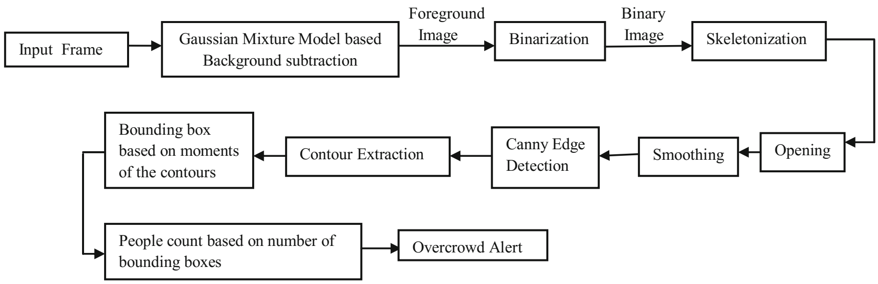


Fig. 11 Architecture of people counting based on skeletonization

Fig. 12 **a** Skeleton. **b** Skeletonization-based algorithm on Mall dataset

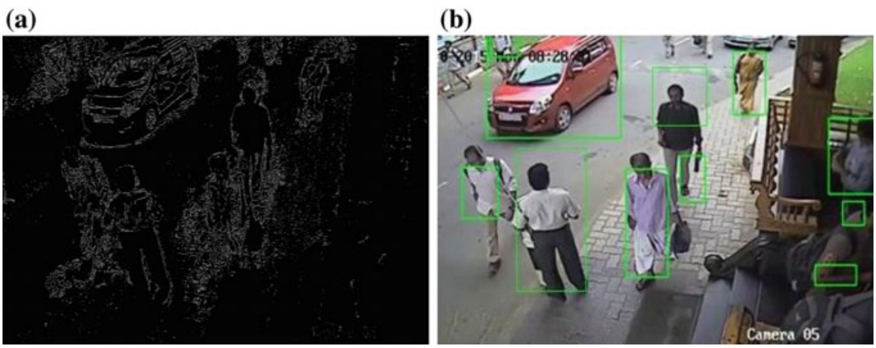


Fig. 13 **a** Skeleton. **b** Skeletonization-based algorithm applied on 223rd frame of the video from surveillance camera of our campus

4 Experimental Results

Open source Computer Vision with C++ has been used for implementing the over-crowd detection algorithms. All the algorithms were tested on two different datasets, Mall dataset [16, 17] and the surveillance dataset from our campus. Each dataset has videos captured under different illumination conditions and with varying concentration and occlusion of people in it.

All the algorithms were tested on 35 videos in Mall Dataset and 40 videos in surveillance dataset from our campus. The videos ranged from 100 frames to 6304 frames. For a video from mall dataset with 100 frames, based on the ground truth of each frame and the number of people identified by each algorithm in each frame, the graphs plotted are shown in Fig. 14a–d.

Graphs show that compared to HoG with SVM, Haar Cascade, and Morphological Operator based algorithms, the proposed algorithm gives people count very close to the ground truth which is calculated manually.

Based on the confusion matrix which depicts the true positive, false positive and false negative, the precision and recall are calculated. True positive stands for people being detected correctly, false positive where people are detected erroneously, false negative indicates non-detection of people by the algorithm in spite of their presence.

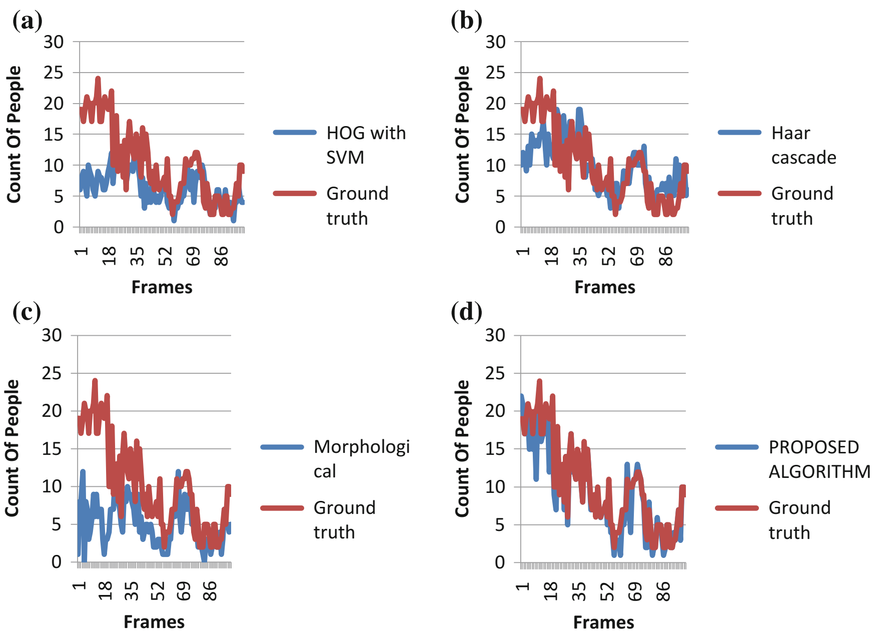


Fig. 14 **a** HoG with SVM. **b** Haar cascade. **c** Based on morphological operator. **d** Proposed algorithm

Table 1 Confusion matrix for Mall dataset video with 600 frames (Ground truth: 6186)

Mall dataset	Algorithm	True positive	False positive	False negative
	HOG with SVM	1602	204	2088
	Haar cascade	3834	842	1218
	Morphological operator based algorithm	222	43	3204
	Proposed algorithm	5142	84	438

Table 2 Confusion matrix for a video from surveillance camera with 812 frames (Ground truth: 8203)

Dataset from the surveillance camera of our campus	Algorithm	True positive	False positive	False negative
	HOG with SVM	2143	271	2684
	Haar cascade	5105	1015	1724
	Morphological operator based algorithm	305	62	4282
	Proposed algorithm	6976	114	574

True negative is not considered for calculating precision and recall as true negative is an indication of people not being detected when people are not there, which is not a prominent indication required for counting the number of people.

Table 1 indicates the confusion matrix for Mall dataset tested for HoG with SVM, Haar cascade, Morphological based, and the proposed Skeletonization-based people counting algorithms. A video which comprised of 600 frames was considered and the ground truth was calculated manually from each frame of the video which counted up to 6186 people.

Table 2 is the confusion matrix for the video from the surveillance camera of our campus tested for HoG with SVM, Haar cascade, Morphological-based, and the proposed Skeletonization-based people counting algorithms. A video which consisted of 812 frames was considered and the ground truth was calculated manually from each frame of the video which counted up to 8203 people.

Table 3 is generated based on the values from Tables 1 and 4 are generated based on the values from Table 2. Precision, recall, and accuracy are calculated. Precision is calculated as

$$\text{Precision} = \frac{\text{True positive}}{\text{True Positive} + \text{False positive}}$$

Table 3 Precision, recall, and accuracy for the confusion matrix in Table 1

Mall dataset	Algorithm	Precision	Recall	Accuracy
	HOG with SVM	0.887	0.434	0.411
	Haar cascade	0.819	0.75	0.65
	Morphological operator based algorithm	0.83	0.065	0.064
	Proposed algorithm	0.984	0.922	0.907

Table 4 Precision, recall, and accuracy for the confusion matrix in Table 2

Dataset from the surveillance camera of our campus	Algorithm	Precision	Recall	Accuracy
	HOG with SVM	0.888	0.444	0.42
	Haar cascade	0.83	0.747	0.66
	Morphological operator based algorithm	0.83	0.07	0.071
	Proposed algorithm	0.99	0.924	0.91

Recall is calculated as

$$\text{Recall} = \frac{\text{True Positive}}{\text{True positive} + \text{False Negative}}$$

Accuracy is calculated as

$$\text{Accuracy} = \frac{\text{True Positive}}{\text{True positive} + \text{False positive} + \text{False Negative}}$$

From the metrics evaluated it is observed that for both the datasets the proposed skeletonization based algorithm gave better precision, recall, and accuracy compared to the other three algorithms. Even though HoG with SVM, Haar cascade and morphological operator based algorithms seemed to provide better precision but failed to provide good accuracy due to the increase in the false negatives reported. Hence the proposed algorithm outperformed for all the other three algorithms on all the varied videos with which it was tested.

5 Conclusion

Crowd alert based on people counting was done using Skeletonization-based algorithm and has been compared with three algorithms like HoG with SVM, Haar Cascade, and Morphological-based algorithms. The confusion matrix and the graphical plot prove that the proposed skeletonization-based algorithm works better than the other algorithms tested upon. The test was done on varied datasets of different sizes and under varied illumination conditions and varied concentration and occlusion of people. Future scope of this work is to improvise on occlusions which are complex.

Acknowledgements The proposed work is tested on three different datasets. (i) Mall dataset [16, 17]; (ii) Dataset from surveillance camera of Amrita Vishwa Vidyapeetham, the university with which we work and where the proposed work is carried on; (iii) User defined video, which is video of myself on which the proposed work was tested upon.

References

1. Sağun MAK, Bolat B (2017) A novel approach for people counting and tracking from crowd video. In: IEEE international conference on INnovations in Intelligent SysTems and Applications (INISTA). IEEE, pp 277–281
2. Hassan MA, Pardiansyah I, Malik AS, Faye I, Rasheed W (2016) Enhanced people counting system based head-shoulder detection in dense crowd scenario. In: 6th International conference on intelligent and advanced systems (ICIAS). IEEE, pp 1–6
3. Neethu A, Bijlani K (2016) People count estimation using hybrid face detection method. In: International conference on information science (ICIS). IEEE, pp 144–148
4. Li B, Zhang J, Zhang Z, Xu Y (2014) A people counting method based on head detection and tracking. In: International conference on smart computing (SMARTCOMP). IEEE, pp 136–141
5. Wang Y, Lian H, Chen P, Lu Z (2014) Counting people with support vector regression. In: 10th International conference on natural computation (ICNC). IEEE, pp 139–143
6. Gao L, Wang Y, Wang J (2016) People counting with block histogram features and network flow constraints. In: International congress on image and signal processing, biomedical engineering and informatics (CISP-BMEI). IEEE, pp 515–520
7. Chen Z, Yuan W, Yang M, Wang C, Wang B (2016) SVM based people counting method in the corridor scene using a single-layer laser scanner. In: 19th International conference on intelligent transportation systems (ITSC). IEEE, pp 2632–2637
8. Perng JW, Wang TY, Hsu YW, Wu BF (2016) The design and implementation of a vision-based people counting system in buses. In: International conference on system science and engineering (ICSSE). IEEE, pp 1–3
9. Baozhu Z, Qiuyu Z, Yufeng X (2016) People counting system based on improved Gaussian background model. In: International conference on smart and sustainable city and big data (ICSSC). IEEE, pp 118–122
10. Riachi S, Karam W, Greige H (2014) An improved real-time method for counting people in crowded scenes based on a statistical approach. In: 11th International conference on informatics in control, automation and robotics (ICINCO), vol 2. IEEE, pp 203–212
11. Cai Z, Yu ZL, Liu H, Zhang K (2014) Counting people in crowded scenes by video analyzing. In: 9th Conference on industrial electronics and applications (ICIEA). IEEE, pp 1841–1845
12. Sikha OK, Sachin Kumar S, Soman KP (2017) Salient region detection and object segmentation in color images using dynamic mode decomposition. J Comput Sci 25:351–366

13. Pranav V, Manjusha R, Parameswaran L (2018) Design of an algorithm for people identification using facial descriptors. Lecture Notes in computational vision and biomechanics, vol 28. Springer, pp 1117–1128
14. Dalal N, Triggs B (2005) Histograms of oriented gradients for human detection. In: IEEE Computer society conference on computer vision and pattern recognition, CVPR, vol 1. IEEE, pp 886–893
15. Viola P, Jones M (2001) Rapid object detection using a boosted cascade of simple features. In: Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition, CVPR, vol 1. IEEE, pp I–I
16. Loy CC, Chen K, Gong S, Xiang T (2011) Feature mining for localised crowd counting. In: British machine vision conference (BMVC), vol 1, no 2
17. Loy CC, Chen K, Gong S, Xiang T (2013) Crowd counting and profiling: methodology and evaluation. In: Modeling, simulation and visual analysis of crowds. Springer, pp 347–382

Performance Comparison of Pre-trained Deep Neural Networks for Automated Glaucoma Detection

Manas Sushil, G. Suguna, R. Lavanya and M. Nirmala Devi

Abstract This paper addresses automated glaucoma detection system using pre-trained convolutional neural networks (CNNs). CNNs, a class of deep neural networks (DNNs), extract features of high-level abstractions from the fundus images, thereby eliminating the need for hand-crafted features which are prone to inaccuracies in segmenting landmark regions and require excessive involvement of experts for annotating these landmarks. This work investigates the applicability of pre-trained CNNs for glaucoma diagnosis, which is preferred when the dataset size is small. Further, pre-trained networks have the advantage of the quick model building. The proposed system has been validated on the High-Resolution (HRF), which is a publicly available benchmark database. Results demonstrate that among other pre-trained CNNs, VGG16 network is more suitable for glaucoma diagnosis.

1 Introduction

Many ocular diseases progress asymptomatically, ultimately resulting in vision impairment [1, 2]. Glaucoma is a prominent cause of blindness, with noticeable vision loss occurring only in advanced stages. It has no cure but early diagnosis can slow progression rate. Worldwide, glaucoma count in the age group of 40–80 years was estimated to be 64.3 million in 2013 and is projected as 76 million in 2020 [3]. Glaucoma affects the optic nerve and is associated with the improper draining

M. Sushil · G. Suguna · R. Lavanya (✉) · M. Nirmala Devi
Department of Electronics and Communication Engineering, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Coimbatore, India
e-mail: r_lavanya@cb.amrita.edu

M. Sushil
e-mail: 66manas@gmail.com

G. Suguna
e-mail: g_suguna@cb.amrita.edu

M. Nirmala Devi
e-mail: m_nirmala@cb.amrita.edu

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_62

of aqueous humor through the drainage angle in the eye. Symptoms include high intraocular pressure (IOP), degeneration of optic nerve fibers, and irreversible vision impairment. Glaucoma diagnosis involves assessing medical history, visual field loss, IOP, as well as Optical Nerve Head (ONH) defects and Retinal Nerve Fiber Layer (RNFL) defects by analyzing retinal images [4].

Digital Fundus Photography (DFP) is a retinal imaging modality that is most widely used in clinical practice for glaucoma screening. It is a quick, simple, cheap, and non-invasive technique that captures a large retinal field. However, manual diagnosis using fundus is subject to ambiguity and is time-consuming.

The advances in image processing and machine learning have paved way for automated diagnosis systems, which can aid time-effective and objective diagnosis. Many researchers have proposed automated systems for diagnosing glaucoma in fundus images. The approaches employed are classified as segmentation-based and non-segmentation-based techniques. In the former approach, also called structural analysis, retinal structures of interest are detected to extract clinically relevant features which ophthalmologists normally use for diagnosis. On the contrary, the non-segmentation-based approach is used to extract high level, non-clinical information from the images, which normally cannot be perceived by human experts [5].

2 Related Work

The segmentation approach in automated glaucoma diagnosis involves localization and extraction of anatomical structures for analyzing defects in the ONH and RNFL. The clinical features of importance that characterize the ONH defects include the Cup-to-Disc Ratio (CDR), ISNT rule and notching. Among these parameters, CDR is widely used in clinical practice; nevertheless, it is susceptible to large inter-individual variability. RNFL defects serve as an early indicator of Glaucoma. However, the detection differentiating RNFL striations and vasculature structure is quite challenging.

Major steps involved in automated glaucoma diagnosis for determination of CDR include pre-processing of the fundus image, ONH localization, OD segmentation, and OC segmentation. Nayak et al. [6] employed neural network for classifying the images as normal or glaucoma based on CDR. An accuracy of 90% was obtained for glaucoma diagnosis. The accuracies of OD and OC segmentation in fundus images face several challenges and have a cascaded effect on the accuracy of glaucoma diagnosis. The segmentation accuracy can be affected by many factors that include low contrast and blood vessel occlusion. Little work has been reported on the analysis of RNFL defects using 2-D retinal images, demonstrating preliminary results.

Some approaches using non-clinical features that are independent of segmenting the retinal structures, such as Textural analysis, Fourier and Wavelet transforms have also been proposed for glaucoma diagnosis [7, 8]. Chen et al. [9] employed convolutional neural network (CNN) to capture hidden, discriminative patterns related to glaucoma resulting in AUC of 0.887.

3 Deep Neural Networks

The conventional segmentation approach based on pre-defined hand-crafted clinical features is not optimal, as the characteristics of the underlying problem may be patient specific. Recent developments in the field of deep learning (DL) offer powerful tools to intelligent image analysis. DL is a promising machine learning technique that is capable of characterizing hidden complex features by applying neural network architectures with multiple hidden layers to solve complex problems.

Deep neural networks (DNN) exhibit feature hierarchy and learn features of increasing complexity and abstraction with increasing network depth. A DNN can discover latent structures within the raw data automatically, eliminating the need for manual labor and high-quality expert knowledge to generate hand-crafted features. Further, it is independent of the cascading effect of segmentation inaccuracy associated with extraction of hand-crafted features [10].

CNN is a branch of deep learning that can visualize and classify visual imagery and is being successfully employed to tackle complex image recognition tasks. CNNs have demonstrated to perform well for medical image analysis and interpretation [9, 11].

4 Methodology

The proposed methodology for automated glaucoma diagnosis using deep learning is illustrated in Fig. 1.

Various steps involved in the proposed work are explained below.

4.1 Data Augmentation

This is a strategy used to increase the images in the dataset. This is performed by combinations of flipping the images with 90° and 270° rotation, left–right rotation, and top–bottom rotation.

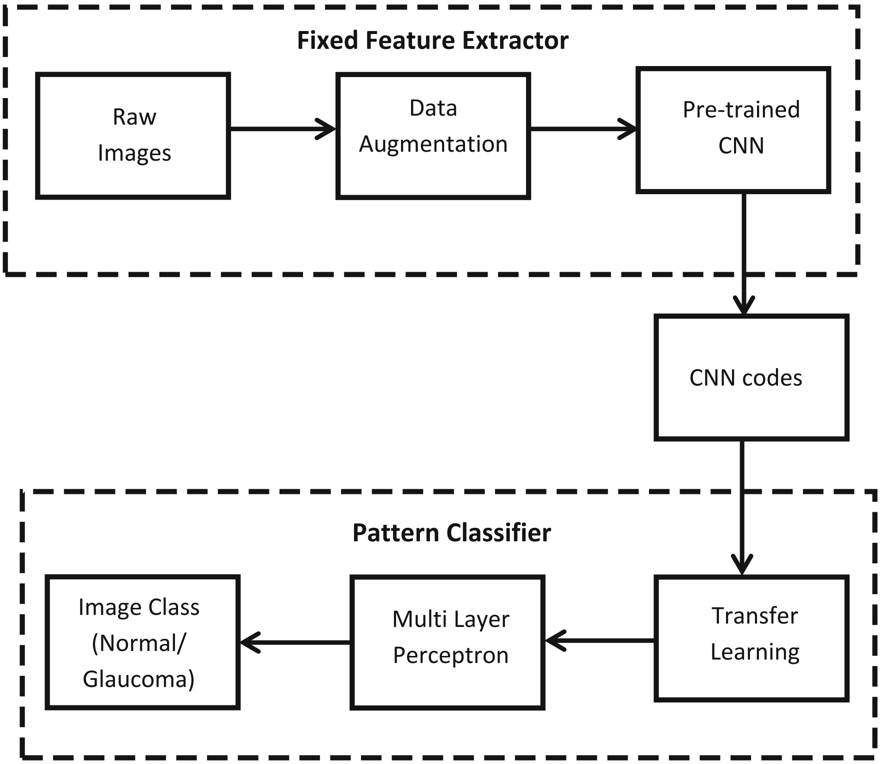


Fig. 1 Proposed methodology

4.2 *Pre-trained CNN*

In practice, random initialization for training a CNN with is rare as a sufficiently large dataset is not available in most cases. Instead, a CNN pre-trained on large datasets are used as an initial network or a fixed feature extractor. This approach is called transfer learning.

In this work, VGG16 architecture [12], a network pre-trained on ImageNet [13] has been employed. ImageNet is a large image database containing 1.2 million images with 1000 categories. VGG16 is a very deep convolution neural network containing 16–19 weight layers as shown in Table 1. Due to this, VGG16 has shown significant improvements over other networks as reported in many studies. At ImageNet challenge 2014, VGG16 secured the first place in localization task and the second place in the classification task.

Table 1 Configuration of VGG16 net[a]

VGG Configuration		
16 weight layers	16 weight layer	19 weight layer
C	D	E
Input image		
CL3-64 cnn3-64	CL3-64 CL 3-64	CL3-64 CL3-64
Maxpool		
CL3-128 CL3-128	CL3-128 CL3-128	CL3-128 CL3-128
Maxpool		
CL3-256 CL3-256 CL1-256	CL3-256 CL3 -256 CL3-256	CL3-256 CL3-256 CL3-256 CL3-256
Maxpool		
CL3-512 CL3-512 CL1-512	CL3-512 CL3-512 CL3-512	CL3-512 CL3-512 CL3-512 CL3-512
Maxpool		
CL3-512 CL3-512 CL1-512	CL3-512 CL3-512 CL3-512	CL3-512 CL3-512 CL3-512 CL3-512
Maxpool		
FCL-4096		
FCL-4096		
FCL-1000		
Soft-max		

[a]Covolutional layer parameters—"CL<filter size>-<number of filters>"
FCL fully connected layer

4.3 Transfer Learning

Two different approaches are widely employed for transfer learning. Generally, when large datasets are involved, the weights of all the layers of pre-trained network layers are fine-tuned using back propagation, as there is no concern of overfitting. However, for small datasets, such as the one employed in this work, fine-tuning will result in overfitting. As an alternate strategy, the final layer of the network is removed, with the other layers serving as a fixed feature extractor. These features are called CNN codes. These codes are used to train a much simpler classifier using the data set under consideration.

5 Results

The performance of pre-trained CNNs in diagnosing glaucoma is evaluated using High-Resolution Fundus (HRF) [14] database which comprises 15 normal fundus

Table 2 Comparison of the performance of pre-trained networks

	VGG16	InceptionV3	ResNet50
Loss	0.07	0.4	0.4
Precision	1	0.48	0.4
Recall	1	0.48	0.08
Accuracy	1	0.6	0.6

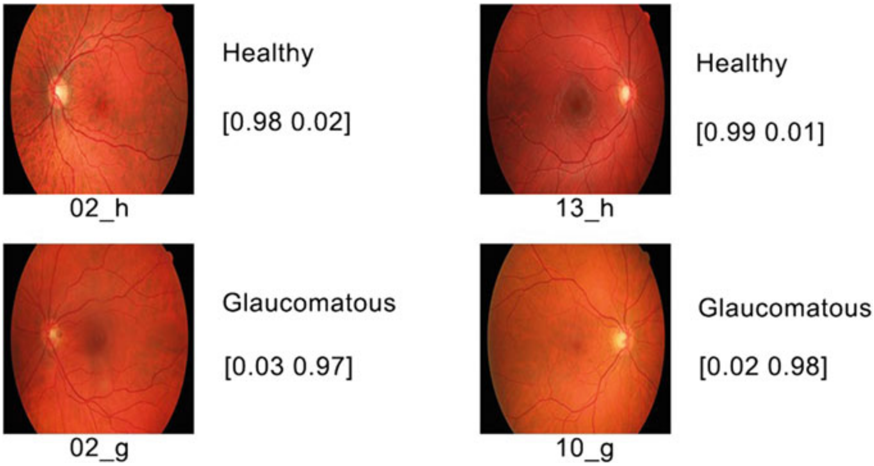


Fig. 2 Softmax outputs of sample HRF images

images and an equal number of glaucomatous patients. Data augmentation was performed to increase the number images to 120, comprising 60 healthy cases and 60 glaucoma cases. Thirty images corresponding to healthy cases and glaucoma cases each used for training. A similar distribution was used for testing. Transfer learning was performed with VGG16, and the results were compared with other pre-trained CNNs including InceptionV3 and ResNet50. As HRF is a relatively small database, the pre-trained networks were used as fixed feature extractors with fully connected layers being replaced by a Multilayer Perceptron (MLP). The MLP employed consists of one hidden layer followed by the output layer. The hidden layer contains 1024 neurons with "Relu" activation and a dropout of 0.5 to avoid overfitting. Output layer consists of two neurons with "Softmax" activation. Fundus image of glaucomatous patients and healthy patients were labeled as [0 1] and [1 0], respectively. Performance of the networks was measured using mean squared error.

In Table 2, the performance of the three pre-trained CNNs is compared. It is observed that VGG16 yields a categorical accuracy of 100% when compared to 60% accuracy achieved by the other two networks. Further, the loss associated with VGG16 is trivial and equal to 0.07, while it is 0.4 each for the other two networks. Sample images from the HRF database with the respective softmax outputs of VGG16 are depicted in Fig. 2.

6 Conclusion

The suitability of pre-trained CNNs for automated glaucoma diagnosis has been investigated in this work. Promising results have been achieved with VGG16 architecture for the High-Resolution Fundus (HRF) database. Future work involves the extension of this work to larger databases using fine-tuning approach for transfer learning.

References

1. Dharani V, Lavanya R (2017) Improved microaneurysm detection in fundus images for diagnosis of diabetic retinopathy. In: International symposium on signal processing and intelligent recognition systems. Springer, pp 185–198 (2017)
2. Sharma A, Subramaniam SD, Ramachandran KI, Lakshmikanthan C, Krishna S, Sundaramoorthy SK (2016) Smartphone-based fundus camera device (MII Ret Cam) and technique with ability to image peripheral retina. Eur J Ophthalmol 26(2):142–144
3. Richter GM, Anne LC (2016) Minimally invasive glaucoma surgery: current status and future prospects. Clin Ophthalmol (Auckland, NZ) 10:189–206
4. Zhang Z, Srivastava R, Liu H, Chen X, Duan D, Wong WK, Kwoh CK, Wong TY, Liu Y (2014) A survey on computer aided diagnosis for ocular diseases. BMC Med Inf Decis Making 14(1):80
5. Haleem MS, Han L, van Hemert J, Li B (2013) Automatic extraction of retinal features from colour retinal images for glaucoma diagnosis: a review. Comput Med Imaging Graph 37(7):581–596
6. Nayak J, Acharya R, Bhat PS, Shetty N, Lim TC (2009) Automated diagnosis of glaucoma using digital fundus images. J Med Syst 33(5):337
7. Bock R, Meier J, Nyúl LG, Hornegger J, Michelson G (2010) Glaucoma risk index: automated glaucoma detection from color fundus images. Med Image Anal 14(3):471–481
8. Dua S, Acharya UR, Chowriappa P, Sree SV (2010) Wavelet-based energy features for glaucomatous image classification. IEEE Trans Inf Technol Biomed 16(1):80–87
9. Chen X, Xu Y, Wong DWK, Wong TY, Liu J (2015) Glaucoma detection based on deep convolutional neural network. IEEE EMBC 2015:715–718
10. Schmidhuber J (2015) Deep learning in neural networks: an overview. Neural Networks 61:85–117
11. Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, Kim R (2016) Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. JAMA 316(22):2402–2410
12. Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556
13. Deng J, Dong W, Socher R, Li LJ, Li K, Fei-Fei L (2015) Imagenet: a large-scale hierarchical image database. IEEE CVPR 2009:248–255
14. Odstrcilik J, Kolar R, Budai A, Hornegger J, Jan J, Gazarek J, Angelopoulou E (2013) Retinal vessel segmentation by improved matched filtering: evaluation on a new high-resolution fundus image database. IET Image Proc 7(4):373–383

Investigation on Land Cover Mapping of Large RS Imagery Using Fuzzy Based Maximum Likelihood Classifier

B. R. Shivakumar and S. V. Rajashekararadhya

Abstract The success of a large number of real-world applications such as mapping, forestry, and change detection depends on the effectiveness with which land cover classes are extracted from Remotely Sensed (RS) imagery. Application of Fuzzy theory in remote sensing has been of great interest in the remote sensing fraternity particularly when the data are inherently Fuzzy. In this paper, a Fuzzy theory based Maximum Likelihood Classifier (MLC) is discussed. The study aims at amplifying the classification accuracy of large heterogeneous multispectral remote sensor data characterized by the overlapping of spectral classes and mixed pixels. Landsat 8 multispectral data of North Canara District was collected from USGS website and is considered for the research. Seven land use land cover classes were identified over the study area. The study also aims at achieving classification results with a confidence level of 95% with ±4% error margin. The conducted research attains the predicted classification accuracy and proves to be a valuable technique for classification of large heterogeneous RS multispectral imagery.

1 Introduction

Classification of multispectral imagery has formed itself as one of the most sought-after technique for information extraction. The process of image classification, in the context or remote sensing, has been broadly classified into hard and soft classification techniques. In hard classification, a pixel is assumed to be an indecomposable part of the image and belongs to just one of the defined land cover class. However, in the real world due to the presence of mixed pixels (mixels), which form the salient feature of heterogeneous study areas, hard classifiers are shown to produce poor

B. R. Shivakumar (✉)
NMAM Institute of Technology, Nitte, India
e-mail: shivkumarbr@gmail.com

S. V. Rajashekararadhya
Kalpataru Institute of Technology, Tiptur, India
e-mail: svraradhya@gmail.com

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_63

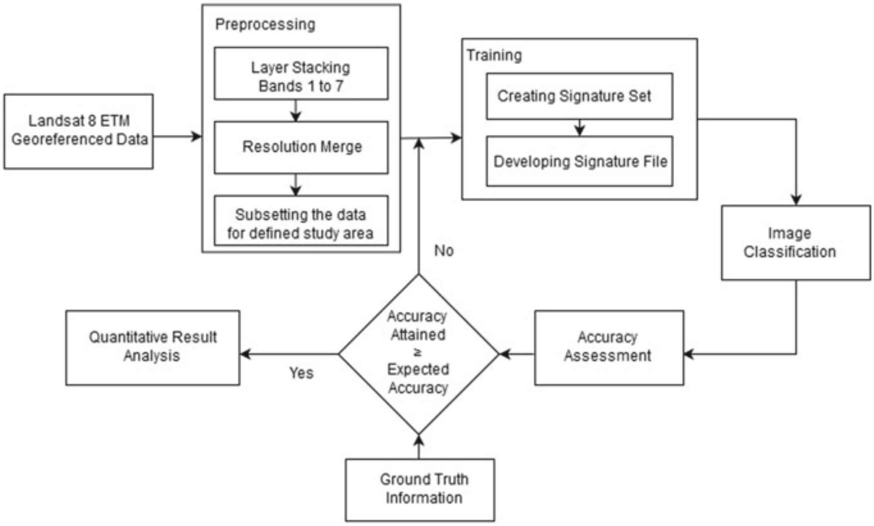


Fig. 1 General flow of the methodology

results as the ground data is imprecise [1–3]. Soft or Fuzzy classification comes in handy under these eventualities and has been known for its capability to extract a lot of helpful data from heterogeneous RS data [2, 4]. A large number of studies have been conducted for classifying RS data using Fuzzy logic [2, 5–8]. Though a large number of researchers have conferred their study towards remotely sensed image classification using different classifiers, it has still remained a challenging task within the remote sensing fraternity [1, 9–12]. Hence, it has become necessary to research and produce new classification techniques for obtaining better classification results.

Zadeh's concept of Fuzzy set theory has provided some very useful options for operating with heterogeneous datasets [13]. Using Fuzzy set concepts a pixel in an image is described with a membership function that links the pixel a real number from 0 to 1. This membership function is treated as the probability of the pixel belonging to each class [1, 13]. This principle is shown to provide an efficient solution for mixed pixel issue [3, 8].

The objective of this study is to perform LULC mapping of the considered study area using Fuzzy theory based MLC. The goal of classification is to obtain results with a confidence level of 95% with ±4% error margin. The methodology followed is illustrated in Fig. 1.

The rest of the paper is organized as follows. Section 2 discusses the data and study area used. Section 3 discusses the classification method employed. Section 4 presents the results obtained during the study. Section 5 presents the conclusions drawn from the results.

2 Materials and Methodology

Land Satellite (Landsat) 8 data was accessed by U.S. Geological Survey (USGS) website. Figure 2 indicates the study area considered and it envelops the North Canara District in Karnataka, India. The Western Ghats form the main geographic feature of the region and runs from North to South. The average rainfall on the coastal part is 3000 mm (120 in) and is as high as 5000 mm (200 in) in the west-facing slopes of Western Ghats. East facing ridge of the Western Ghats is the rain shadow region and receives, on an average, only 1000 mm (39 in) rainfall annually [14]. High rainfall in the region supports lavish forests that coat over 70% of the coverage area. The north part of the Western Ghats forms Moist Deciduous Forests ranging from 250 to 1000 m in elevation. Above 1000 m elevation are the Evergreen rain forests [15]. The study area also has chunks of degraded scrub jungles and savanna. The beach region is characterized by coconut plantations and screw pine. The study area is recognized as a coastal agro-climatic zone by the government of India. The overall spatial resolution of the image is 15 m. This data was acquired on May 18 2016, which is the mid-summer season and is free from clouds. Layer stacking of first seven bands produces a low-resolution multispectral (MS) imagery of 30 m. Resolution merge technique is used to merge the 30 m MS image with higher spatial resolution band 8 to obtain high resolution 15 m multispectral image. This image is then used for further processing.

2.1 Feature Extraction/Signature Collection

There are several methods of collecting the training data, including; in situ data collection, on-screen selection of polygonal data, and/or on-screen seeding of training data [1]. By employing an on-screen selection of polygonal data technique, seven land use land cover classes were identified over the study area; (i) Water Body, (ii) Kharif, (iii) Built Up, (iv) Scrub Land, (v) Double Crop/Horticultural Plantations, (vi) Moist Deciduous Forest, and (vii) Evergreen Forest. Spectral signatures were collected for each class with at least 300 sample pixels per class. The study considered both pure and mixed pixels for training the algorithm.

The considered study area exhibits the characteristics of a heterogeneous dataset with more than two classes severely overlapping one another. Class separability was measured in terms of Euclidean distance considering two classes at a time. The Euclidean distance was redefined to measure the severity of land cover overlapping and is used as the spectral similarity index (SSI). Highest class separability was measured between Water Body and Kharif classes and was considered as the reference for other classes. Classes with a SSI value of less than 0.4 were determined to be severely overlapping one another at certain locations. SSI measurements for all possible class pairs are shown in Table 1.

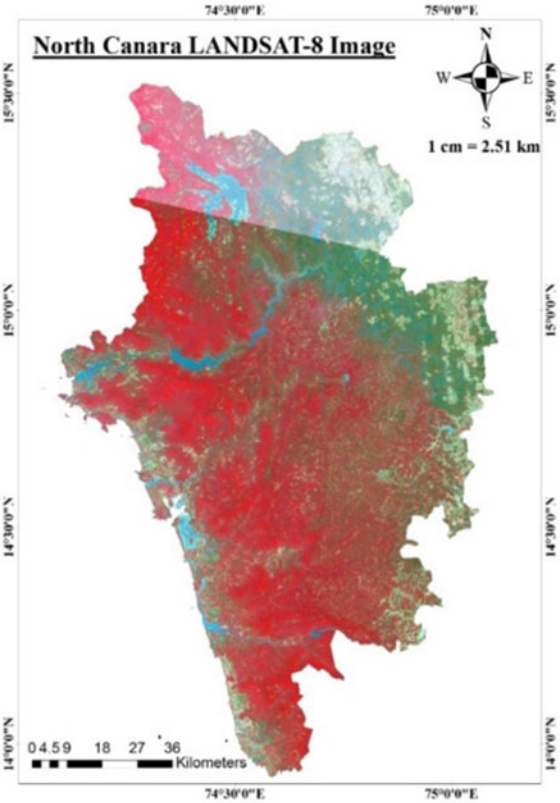


Fig. 2 15 m spatial resolution map of the study area [16]

3 Fuzzy Theory Based Maximum Likelihood Classifier

The Fuzzy theory based ML classifier discussed in this paper works in three stages; (i) Fuzzification of the data, (ii) Classification using Fuzzy theory based MLC, and (iii) Defuzzification.

3.1 Fuzzification

In the Fuzzification step, each pixel on the input image is converted into a pixel measurement vector, x, of membership grades. The Fuzzy membership function for any x must lie in the range 0–1, they should all add up to unity, and should be positive values. These characteristics are listed in (1), (2), and (3) [1].

Table 1 Class separability in terms of spectral similarity index

Class Pair	SSI	Class pair	SSI	Class pair	SSI
1:2	1	2:4	0.441	3:7	0.963
1:3	0.879	2:5	0.796	4:5	0.544
1:4	0.628	2:6	0.486	4:6	**0.061**
1:5	0.666	2:7	0.968	4:7	0.649
1:6	0.598	3:4	0.528	5:6	0.518
1:7	0.668	3:5	0.774	5:7	**0.210**
2:3	**0.338**	3:6	0.562	6:7	0.612

1: Water Body, 2: Kharif, 3: Built Up, 4: Scrub Land, 5: Double Crop/Horticultural Plantations, 6: Moist Deciduous Forest, 7: Evergreen Forest
Bold characters represent the spectral similarity index for the most severely overlapping class pairs. The class pairs having spectral similarity index (SSI) value of 0.3 and lesser are the most severely overlapping class pairs. It is highlighted to make it easy for the reader to identify the severely overlapping class pairs easily

$$0 \leq f_{F_i}(x) \leq 1 \tag{1}$$

$$\sum_{x \in X} f_{F_i}(x) > 0 \tag{2}$$

$$\sum_{i=1}^{m} f_{F_i}(x) = 1, \tag{3}$$

where, F_i is one of the spectral classes, X represents all pixels in the dataset, x is a pixel measurement vector, m is the number of classes, and f_{F_i} is the membership function of the Fuzzy set $F_i(1 \leq i \leq m)$. All the membership function values are recorded as a Fuzzy partition matrix

$$\begin{bmatrix} f_{F_1}(x_1) & f_{F_1}(x_2) & \ldots & f_{F_1}(x_n) \\ f_2(x_1) & f_{F_2}(x_2) & \ldots & f_{F_2}(x_n) \\ \vdots & \vdots & \ldots & \vdots \\ f_{F_m}(x_1) & f_{F_m}(x_2) & \ldots & f_{F_m}(x_n) \end{bmatrix}, \tag{4}$$

where, n represents the total number of pixels, and x_i is the ith pixel's measurement vector [1].

3.2 Classification

Fuzzy based classification involves the use of Fuzzy mean and covariance matrices. For class c, Fuzzy mean is computed as

$$\mu_c^* = \frac{\sum_{i=1}^{n} f_c(x_i)x_i}{\sum_{i=1}^{n} f_c(x_i)}, \tag{5}$$

where, x_i is a sample pixel measurement vector $(1 \le i \le n)$, f_c is the membership function of class c, and n is the total number of sample pixel measurement vectors [1].

The Fuzzy covariance matrix V_c^* is computed as;

$$V_c^* = \frac{\sum_{i=1}^{n} f_c(x_i)\left(x_i - \mu_c^*\right)\left(x_i - \mu_c^*\right)^T}{\sum_{i=1}^{n} f_c(x_i)} \tag{6}$$

These mean and covariance values replace the conventional mean and covariance matrix in the classical MLC algorithm. This will convert a classical MLC algorithm into a Fuzzy based soft classification algorithm [1].

Fuzzy set theory solely provides membership functions for every pixel over the defined number of classes and requires a parametric rule for assigning those pixels to relevant classes. Parametric rules such as Maximum Likelihood, Mahalanobis Distance et al. may be used in the process. This study involves the utilization of Maximum Likelihood classifier as the parametric rule.

Fuzzy based Maximum Likelihood classifier uses Fuzzy mean and Covariance matrices replacing the conventional mean and covariance matrices. For an n-band multispectral image the likelihood function for a pixel belonging to class k is given by [1],

$$p^*(x|w_i) = \frac{1}{(2\pi)^{n/2}\left|V_k^*\right|^{1/2}} \exp\left[-\frac{1}{2}\left(x - \mu_k^*\right)V_k^{*-1}\left(x - \mu_k^*\right)^{\mathrm{T}}\right], \tag{7}$$

where, x is one of the brightness values on the x-axis, μ_k^* is the Fuzzy mean as in (5), V_k^* is the Fuzzy covariance matrix as in (6), and $\left(x - \mu_k^*\right)^{\mathrm{T}}$ is the transpose of vector $\left(x - \mu_k^*\right)$. Similarly, $p^*(x|w_i)$ is calculated for each pixel for all classes. A membership function then enables the algorithm to decide to which class the corresponding pixel is to be assigned. For Maximum Likelihood classifier, the membership function can be defined as [1];

$$f_c(x) = \frac{p^*(x|w_k)}{\sum_{i=1}^{m} p^*(x|w_k)} \tag{8}$$

The membership grades of a pixel vector depend on x's position in the vector space. $f_c(x)$ increases exponentially with the decrease of $\left(x - \mu_k^*\right)V_k^{*-1}\left(x - \mu_k^*\right)^{\mathrm{T}}$, i.e., the Mahalanobis distance between x and class k. The factor $\sum_{i=1}^{m} p^*(x|w_k)$ is a normalization factor [1]. Applying this type of Fuzzy logic creates a membership grade matrix for each pixel.

3.3 Defuzzification Using Fuzzy Convolution

Fuzzy convolution technique is used to convert the n-layer output of classification into a map like structure. It creates the map by computing the total weighted inverse distance of all the classes in a window of pixels. The process first computes the total inverse distance summed over the entire set of Fuzzy classification layers for each class. It then assigns the center pixel to the class for which the value $T[k]$ is largest. The total inverse distance, $T[k]$, can be computed using [17]:

$$T[k] = \sum_{i=0}^{s} \sum_{j=0}^{s} \sum_{l=0}^{n} \frac{w_{ij}}{D_{ijl}[k]} \tag{9}$$

where, i is the row index of window, j is the column index of window, s is the size of window, l is the layer index of fuzzy layers used, n is the number of fuzzy layers used, W is the weight table for window, k is the class value, $D[k]$ is the distance file value for class k, and $T[k]$ is the total weighted distance of window for class k [17]. This study considers a 5×5 size window given by

$$\begin{bmatrix} 0.500 & 0.605 & 0.646 & 0.605 & 0.500 \\ 0.605 & 0.750 & 0.823 & 0.750 & 0.605 \\ 0.646 & 0.823 & 1.000 & 0.823 & 0.646 \\ 0.605 & 0.750 & 0.823 & 0.750 & 0.605 \\ 0.500 & 0.605 & 0.646 & 0.605 & 0.500 \end{bmatrix} \tag{10}$$

4 Results and Discussion

Figure 3 illustrates the Fuzzy topology based maximum likelihood classification map of the study area. Table 2 shows the results obtained after accuracy assessment. User's Accuracy and Kappa value are considered as the pivotal parameters in judging the classification process [18]. The classifier extracted dominant classes (Evergreen Forest and Deciduous Forest) very efficiently. Less dominant classes (Double Crop and Built Up) are extracted very poorly by the classifier. Fuzzy topology based maximum likelihood classifier has shown significant improvement in classification performance [12]. An overall Kappa value of 0.7870 indicates an excellent performance from the classifier [18].

To illustrate the usefulness of using the inverse weighted distance of all classes from the weight windows for pixels for assigning pixels to class values, Table 3 indicates the inverse weighted distance, $T[k]$, for 10 pixels randomly selected from the study area. A hard classified map is created by assigning a pixel to the class for which the distance measure, $T[k]$, is maximum.

Table 2 Results of Fuzzy topology based maximum likelihood classification

Class name	Reference totals	Classified totals	Number correct	Producer's accuracy (%)	User's accuracy (%)	Kappa value (k_{hat})
Evergreen Forest	534	538	502	94.01	93.31	0.8564
Scrub Land	59	70	40	67.80	57.14	0.5446
Moist Deciduous Forest	235	214	205	87.23	95.79	0.9450
Built Up	5	14	3	60.00	21.43	0.2103
Double Crop/Horticultural Plantations	57	80	37	64.91	46.25	0.4300
Water Body	31	16	16	51.61	100.00	1.0000
Kharif	79	68	59	74.68	86.76	0.8563
Total	1000	1000	862			
Overall classification accuracy = 86.2%						
Overall Kappa Statistic = 0.7870						

Table 3 Total inverse weighted distance of all classes for selected pixels

Pixel no.	Weighted inverse distance ($T[k]$) from							Class assigned
	EGF[a]	SL[b]	MDF[c]	BU[d]	DC/HP[e]	WB[f]	KH[g]	
1	**0.1988**	0.1199	0.1159	0.0654	0.1664	0.0026	0.0463	**EGF[a]**
2	0.1243	**0.1714**	0.1593	0.0821	0.1359	0.0011	0.1041	**SL[b]**
3	0.1426	0.1592	**0.1701**	0.0949	0.1348	0.0012	0.0815	**MDF[c]**
4	0.0639	0.1140	0.0576	**0.1655**	0.1443	0.0022	0.1345	**BU[d]**
5	0.0635	0.1561	0.0989	0.1361	**0.1709**	0.0019	0.1431	**DC/HP[e]**
6	0.0468	0.0587	0.0432	0.0951	0.0768	**0.1552**	0.0701	**WB[f]**
7	0.0148	0.1188	0.1634	0.1366	0.0540	0.0011	**0.1748**	**KH[g]**
8	0.0295	0.1161	0.0541	0.1059	0.0920	0.0007	**0.1939**	**KH[g]**
9	0.0215	0.0761	0.0314	**0.1742**	0.0582	0.0025	0.1235	**BU[d]**
10	0.1791	0.0758	0.0568	0.0326	**0.2013**	0.0018	0.0229	**DC/HP[e]**

EGF[a]: Evergreen Forest, SL[b]: Scrub Land, MDF[c]: Moist Deciduous Forest, BU[d]: Built Up, DC/HP[e]: Double Crop/Horticultural Plantations, WB[f]: Water Body, KH[g]: Kharif
Bold characters represent the spectral similarity index for the most severely overlapping class pairs. The class pairs having spectral similarity index (SSI) value of 0.3 and lesser are the most severely overlapping class pairs. It is highlighted to make it easy for the reader to identify the severely overlapping class pairs easily

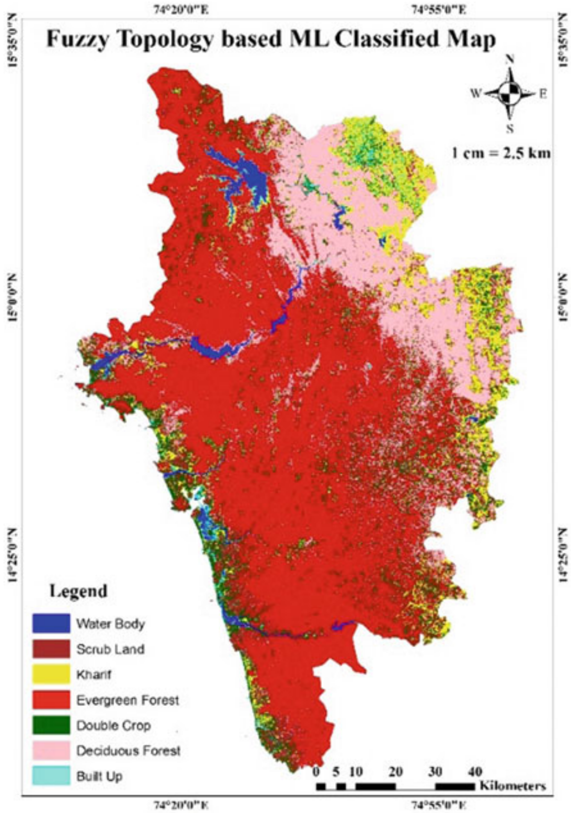


Fig. 3 Fuzzy topology based maximum likelihood classified map of the study area

4.1 *Placing Confidence Limits on Assessed Accuracy*

A straightforward statistical approach may be used to express the interval within which the true map accuracy lies, say, with 95% certainty. It is possible to use the normal distribution to obtain this interval by the expression [2];

$$-Z_{\alpha/2} < \frac{x - nP}{\sqrt{nP(1-P)}} < Z_{\alpha/2}, \quad (11)$$

where, n is the number of testing pixels, $x(=np)$ is the number of correctly labelled pixels, P is the thematic map accuracy in percentage, $p(=x/n)$ is the proportion of pixels correctly classified, $Z_{\alpha/2}$ is the value of the normal distribution beyond which on both tails α of the population is excluded [2].

From Eq. (11), the estimate of the thematic map accuracy, P, estimated by the proportion of pixels that are correctly classified in the testing set, at the 95% confidence level, for large values of n and x, and for reasonable accuracies are [2],

$$\frac{x \pm 1.960\sqrt{\frac{x(n-x)}{n}} + 0.960}{n} = p \pm \frac{1.960}{n}\sqrt{\frac{x(n-x)}{n}} + 0.960 \quad (12)$$

For 1000 testing pixels and a minimum of 80% of expected accuracy, it is expected to have, at least, 800 pixels to be correctly classified. From (12), the bounds on the estimated map accuracy are $P = p \pm 0.039$ or, in percentage terms, the map accuracy is approximated to be between 82.1 and 89.9%.

5 Conclusion

In this paper, a novel method for classification of large RS imagery through embedding Fuzzy theory into classical Maximum Likelihood Classifier (MLC) is discussed. A pixel is the basic building block of an image and is indecomposable in hard classification applications. The study proves that a pixel can be used as a decomposable unit in image classification. The results obtained indicate that Fuzzy theory based MLC permits obtaining accurate results for large heterogeneous study areas in the presence of mixed pixels and spectrally overlapping classes. The study conjointly confirms that one can use mixed pixels as training data and yet achieve good results. The objective of obtaining classification results with confidence level of 95% with ±4% error margin is achieved. Hence, it can be concluded that the discussed Fuzzy topology based MLC handles mixed pixel issue more successfully. However, more investigation is needed on the classification performance of some classes, such as Built Up and Double Crop/Horticultural Plantations. Future scope of the work involves exploring the information richness provided by embedding Fuzzy theory into MLC for other sensor data.

References

1. Jensen JR (2000) Introductory digital image processing: a remote sensing perspective. Prentice-Hall Inc., New Jersey
2. Zhang J, Foody GM (1998) A fuzzy classification of sub-urban land cover from remotely sensed imagery. Int J Remote Sens 19:2721–2738
3. Wang F (1990) Fuzzy supervised classification of remote sensing images. IEEE Trans Geosci Remote Sens 28:194–201
4. Ji M, Jensen JR (1996) Fuzzy training in supervised digital image processing. Geogr Inf Sci 2:1–11
5. Wang Y, Jamshidi M (2004) Fuzzy logic applied in remote sensing image classification. Syst Man Cybern 2004(7):6378–6382
6. Melgani F, Al Hashemy BAR (2000) Taha SMR (2000) An explicit fuzzy supervised classification method for multispectral remote sensing images. IEEE Trans Geosci Remote Sens 38(1):287–295
7. Droj G (2007) The applicability of fuzzy theory in remote sensing image classification. Stud Univ Babes, Bolyai, Inform LII, 89–96

8. Wang F (1990) Improving remote-sensing image-analysis through fuzzy information representation. Photogramm Eng Remote Sens 56:1163–1169
9. Lu D, Weng Q (2007) A survey of image classification methods and techniques for improving classification performance. Int J Remote Sens 28:823–870
10. Mas JF, Flores JJ (2008) The application of artificial neural networks to the analysis of remotely sensed data. Int J Remote Sens 29:617–663
11. Maselli F, Conese C, De Filippis T, Romani M (1995) Integration of ancillary data into a maximum-likelihood classifier with nonparametric priors. ISPRS J Photogramm Remote Sens 50:2–11
12. Shivakumar BR, Rajashekararadhya SV (2017) Spectral similarity for evaluating classification performance of traditional classifiers. In: 2017 International conference on wireless communications signal processing and networking (WiSPNET). Chennai, pp 1999–2004
13. Zadeh LA (1965) Fuzzy sets. Inf Control 8:338–353
14. Directorate of Census Operations Karnataka: District census handbook Uttara Kannada (2014)
15. Pascal JP (1986) Explanatory booklet on forest map of South India: Belgaum-Dharwar-Panaji, Shimoga, Mercara-Mysore. In: Institut Francais de Pondichery, p 88. Travaux de la Section Scientifique et Technique. Hors Serie N 18
16. USGS: Earth explorer, https://earthexplorer.usgs.gov/
17. Pouncey R, Swanson K, Hart K (1999) ERDAS field guide
18. Richards JA (2013) Remote sensing digital image analysis. Springer

Gaussian Membership Function and Type II Fuzzy Sets Based Approach for Edge Enhancement of Malaria Parasites in Microscopic Blood Images

Golla Madhu

Abstract This research presents a three-stage approach. In the first stage, the original image transformed into grayscale image, then normalizes grayscale image using min-max normalization, which performs a linear conversion on the original image data. The second stage calculates the Gaussian membership function on the normalized grayscale image then measure lower membership values and upper membership values using a threshold value. In addition, computed a novel membership function with Hamacher t-conorm using lower and upper membership values on given images. Finally, the median filter applied on these images to obtain edge enhanced microscopic images. The current study is conducted on the microscopic blood images of the malaria parasites. The experimental results compared with Prewitt filter, Sobel edge filter, and rank-ordered filter. The proposed approach is consistent and coherent in all microscopic malaria parasite images with four stages, with average entropy 0.90215 and 59.69% PSNR values, respectively.

1 Introduction

Medical imaging analysis has been increasing rapidly with the advent of imaging technologies and it plays a significant role in diagnosing a disease and for observing patient's health disorders and giving a competent treatment [1]. Most of the medical images such as blood images, magnetic resonance imaging (MRI), ultrasonic scanning, and X-rays consist with low contrast images [2]. In addition, these images are not clear because of their blur and uncertainties in edges or boundaries of the images. To handle these problems, we require a preprocessing technique that obtains the quality and clear edge images. Hence, edge enhancement technique plays a vital role in medical imaging analysis. In addition, it brings out the relevant features, which are not accurately detectable and interrupt redundant information [3].

G. Madhu (✉)
Department of Information Technology, VNR Vignana Jyothi Institute
of Engineering and Technology, Hyderabad 500090, Telangana, India
e-mail: madhu_g@vnrvjiet.in

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_64

Malaria appears worldwide in more than 100 countries in the tropical and subtropical regions, and it becomes one of the prominent infection diseases, with an outcome of more than 2.3 billion deaths every year. In India, 80% of malaria cases occur among 20% of the population. In the year 2017, about 673,474 positive cases were found and of which 439,934 are of Plasmodium falciparum cases, out of a total 89,297,252 blood smears examined, until September 2017 [Source http://nvbdcp.gov.in/malaria3.html]. Therefore, parasite analysis in thick blood smears microscopic images became the gold standard for the diagnosis of malaria. Then, it requires the better edge enhancement technique for representation of blood cell recognition of the microscopic blood images that represents preprocessing step in image processing. Edge enhancement increase or enhance the cell edge or boundaries of the blood image [3]. One of the popular techniques also known as a median filter that preserves the edges or boundaries of the images while computing the variation with respect to the middle pixel component of the filter, then edges are enhanced on the image [3]. Fuzzy sets used to characterize uncertainty in the form of membership function. This research, focus on the image obtained through microscopic blood images of malaria and use these images as input images. Then, generated grayscale image from input image and then computed Gaussian membership function on grayscale image using interval type II fuzzy sets, which represents lower membership values and upper membership values. Using these membership values, a novel membership function is computed by utilizing Hamacher t-conorm functionalities. Then, it applied the median filter on these images to obtain edge enhanced images that significantly represent better quality of microscopic blood images. This technique is useful for medical experts to make easy decisions and predictions for further diagnosis of the patients.

2 Related Works

The aim of edge enhancement technique is to increase the quality of an image from the original image in order to highlight significant features of the image [4]. However, edge enhancement plays a significant role in imaging analysis that improves the edge contrast of an image [5]. In the literature, many researchers proposed and established several edge detection techniques, for example, Kirsch [6], Robert edge detection [7], canny edge detection [8], LOG [9], Sobel [10], and Prewitt [11]. However, this approach uses the notion of spatial variance filter using the local gradient approach on image pixels and search for local maxima to establish the stage edges. An advantage of these approaches is to process the image data, which is comparatively less in time to identify the edges and their positioning that computationally optimized. Nevertheless, they are very sensitive to noise data and missed data or to the false edges detected. Mokhtarian and Suomela [12] proposed an innovative approach for image edge detection using curvature scale-space illustration. Genming and Bouzong [13] developed a 5×5 kernel for finding the edges in an image using a threshold value. Conversely, their limitations are inadaptability of two regions with the fluctuating grayscale image due to their fixed threshold value. Fürhapter et al. [14] demonstrated an optical-based edge contrast enhancement in light microscopic images. Chaira [15] proposed an Attanassov's intuitionistic fuzzy sets for image edge detection. Marr and Hildreth [16, 17] suggested edge detection using a Gaussian filter and this approach has been a very popular technique, an earlier Canny liberated his detection. In this research, initially pointed out that the deviation of image intensity, and it happens at different levels of the image. Santis et al. [18] suggested a statistical approach for edge detection approach using the linear stochastic based signal technique, which obtained from the physical image description.

Fuzzy sets are used to perform analytical and logical approaches using rough approximations that are slightly crisp values which have been hired for edge detection significantly to reduce the uncertainty of the medical applications [19, 20]. Khamy et al. [20] recommended an improved fuzzy Sobel filter for edge detection as well as fuzzy reasoning methods, this method used to discover the image edges. Lu et al. [21] suggested a fuzzy set-based neural network (NN) approach for the edge enhancement then detection by improving the misplaced edges and discarding the false edges. Wu et al. [22] suggested a fast multilevel based fuzzy approach for image edge detection, which perform the robust and accurate detection of edges. Bustince et al. [23], proposed an interval valued-based fuzzy approach for edge detection of the image, then it is applied on common images but not on cellular images. Barrenechea et al. [24] discussed an interval-valued based fuzzy relation for the generation of the fuzzy edges on image data. Chaira [25] proposed an intuitionistic-based fuzzy set are used to edge enhancement of medical image data. Chaira [26] proposed a rank-order-based filter for edge enhancement of cellular images using an interval type II fuzzy sets and whole variation of the image of the edge enhancement that makes cell objects additional bulging with respect to background. Talari Tirupal et al. presented Sugeno's intuitionistic-based fuzzy set on medical image data which converted into Sugeno's intuitionistic based fuzzy image.

3 Interval Type II Fuzzy Sets

This research, presents basic notions of interval type II fuzzy sets and some important mathematical definitions and precise terms that will successfully communicate to proposed works [27]. In addition, interval type II fuzzy sets are differentiating from the type I fuzzy sets via adding uncertainty. Prof. Zadeh (1975) identified there was a difficulty with the type I fuzzy sets when he developed a fuzzy II and higher types [28]. The type I fuzzy set has commonly used in various real-world applications.

However, complex issues usually involve higher degree of uncertainty that cannot handle by the type I fuzzy set. To overcome such an issues, proposed an extension of the type I fuzzy set similarly known as the interval type II fuzzy sets [29]. In the interval-based type I fuzzy set membership, represent a type of vagueness or uncertainties. A sample of the basic Gaussian membership function is shown in Fig. 1a. Therefore, the type I fuzz set A is represented through $\mu_A(x)$, where

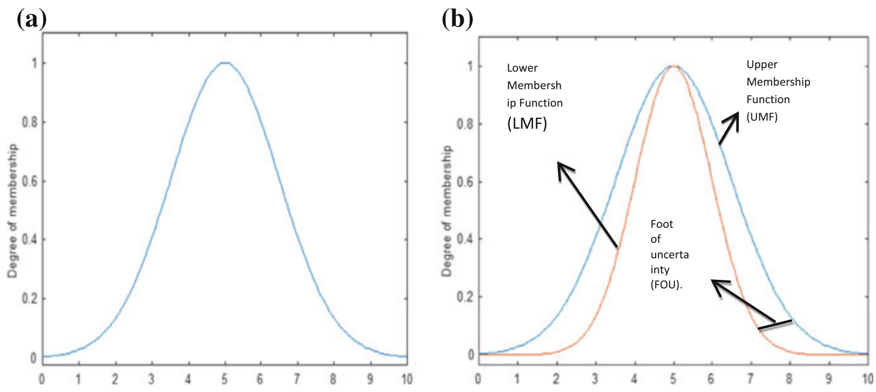


Fig. 1 **a** Represents the type I fuzzy set membership function, **b** Represents the interval type II fuzzy membership function also shown lower membership function (LMF), upper membership function (UMF) and foot of uncertainty (FOU)

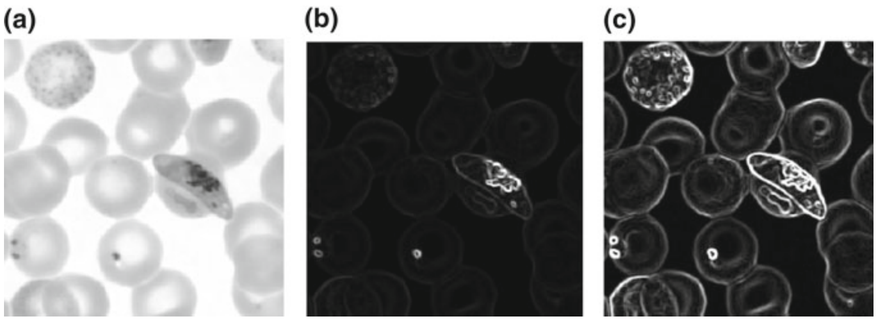


Fig. 2 **a** Representation of blood image with malaria parasite with gametocyte stage, **b** Corresponding edge enhanced results with $\lambda = 0.47$, **c** Edge enhanced results with for $\lambda = 1.37$

$A = \{x, \mu_A(x_{i,j})|x \in X\}$ in [0,1]. Similarly, an interval type II fuzzy sets requires different types of membership grade functions that describe the uncertainty and each membership lies between [0,1]. Interval type II fuzzy set, defined the primary membership function, which represents ambiguity of boundary regions in the form of lower and upper. The uncertainty represented by a region is also known as foot of uncertainty (FOU) shown in Fig. 2. This represents the contour of Interval type II fuzzy set (shown Fig. 2) [30, 31].

Definition The interval type II fuzzy set [31] is defined in terms of membership function, which represents $\mu_{\tilde{A}}(x_{i,j}, a)$ where $x_{i,j} \in X$ and $a \in Y \subseteq [0, 1]$.

$$A = \{((x_{i,j}, a), \mu_{\tilde{A}}(x_{i,j}, a)) | \forall x_{i,j} \in X, \forall a \in Y \subseteq [0, 1] \tag{1}$$

In the Eq. (1), $\mu_{\tilde{A}}(x_{i,j}, a)$ represent the type II membership function and it belongs to $0 \le \mu_{\tilde{A}}(x_{i,j}, a) \le 1$.

To outline the interval type II fuzzy sets, describe an interval type I fuzzy set, and then assign a lower membership and an upper membership function that belongs to each pixel element and to (re)build the FOU [32]. A more concrete description for interval type II fuzzy set represented is as follows:

$$\tilde{A} = \{ (x_{i,j}, \mu_L(x_{i,j}), \mu_U(x_{i,j})) | \text{for all } x_{i,j} \in X, \mu_L(x_{i,j}) \le \mu(x_{i,j}) \le \mu_U(x_{i,j}) \text{ where } \mu(x_{i,j}) \subseteq [0, 1] \tag{2}$$

In the Eq. (2), $\mu_L(x_{i,j})$ and $\mu_U(x_{i,j})$ represents lower membership and upper membership functions, respectively. In our research, these membership functions are represented as follows:

$$\mu_{\text{Lower}}(x_{i,j}) = [\mu(x_{i,j})]^{1/\alpha} \tag{3}$$

$$\mu_{\text{Upper}}(x_{i,j}) = [\mu(x_{i,j})]^{\alpha} \tag{4}$$

Here, the α value belongs to $0 \le \alpha \le 1$, this study, we used $\alpha = 0.37$ because of $\alpha \gg 2$ that is not significant representation for image data [32].

4 Proposed Methodology

The method of edge enhancement is mainly partitioned into three-stage approach that relates the Gaussian membership function using interval type II fuzzy sets. In the first stage, the image is transformed into grayscale image, and then normalized the image using min-max normalization, which performs a linear conversion on the original image data. The second stage calculates the Gaussian membership function on a normalized grayscale image and then measure lower membership values and upper membership values using a threshold value (α). In addition, calculate a novel membership function with Hamacher t-conorm using these lower and upper membership values on microscopic images. Finally, the median filter is applied on these images to obtain edge enhanced microscopic images. This research is conducted on the microscopic blood images of the malaria parasite to enhance the edges of the blood cells. The description of the proposed method is shown in algorithm-1.

In this research, the malaria blood images consist of four different stages like, ring forms, schizonts, trophozoites, and gametocyte stages. Thin peripheral blood smear samples were collected from Vivekananda Hospital, Begumpet, Hyderabad, Telangana 500016, India. Initially, the RGB images are transformed into the grayscale image A with the size M × N. Min-max normalization is to perform the linear transformation on the grayscale image using the following formula.

$$\mu' = \frac{x - \min_A}{\max_A - \min_A}(n\_\max_A - n\_\min_A) + n\_\min_A \tag{5}$$

where x is the gray values of the image that belongs to $[0, L-1]$. $\max_A$ & $\min_A$ are the, maximum and minimum values of the gray level of the image and $n\_\max_A$ & $n\_\min_A$ values are produced from the range of the image based on its class. This normalization formula preserves the relationships among the original image pixel values. Various types of type II fuzzy set based membership functions are proposed and applied in the literature such as Gaussian, triangular, sigmoid and trapezoidal, etc. In this work, we used Gaussian membership function $\mu_{\tilde{A}}(x_{i,j})$ which is defined as follows:

$$\mu_{\tilde{A}}(x_{i,j}) = \exp\left(-\frac{1}{2}\left(\frac{\vec{x}_n - \vec{m}_j}{\sigma_j}\right)^2\right) \tag{6}$$

where $\vec{x}_n = [x_{n1}, x_{n2}, \ldots, x_{nm}]^{\mathrm{T}}$ is vector which is made by the pixel values of the nth sample of image. $\vec{m}_j = \left[x_{j1}, x_{j2}, \ldots, x_{jm}\right]^{\mathrm{T}}$ is mean of the pixel values of the grayscale images denotes as $\vec{m}_j = \frac{1}{N}\sum_{\vec{x}_n \in S_j} \vec{x}_n$ and σ_j standard deviation that takes on values in lower membership and upper membership values. These lower and upper membership values are represented based on the type II fuzzy sets that are calculated using Eqs. (3) and (4). Fuzzy linguistic hedge with $\alpha \leq 1$ and its reciprocal value are drawn to the FOU, our study used $\alpha = 0.37$ that produce the lower membership and upper membership values, respectively. In addition, fuzzy linguistic hedges are utilized to adapt the membership values.

Utilizing the functionalities of generalized "Hamacher t-conorm" [33], we proposed a novel membership function as follows:

$$\mathrm{Mem}(\mu_L(x_{i,j}), \mu_U(x_{i,j})) = \frac{\lambda(\mu_L(x_{i,j}) + \mu_U(x_{i,j})) + \mu_L(x_{i,j}) * \mu_U(x_{i,j})(1-2\lambda)}{(1+\lambda) + \mu_L(x_{i,j}) * \mu_U(x_{i,j})(1-\lambda)} \text{ for } \lambda \geq 0 \tag{7}$$

where $\mu_L(x_{i,j})$, $\mu_U(x_{i,j})$ represents the lower membership values, upper membership values of image dataset that are calculated from Eqs. (1) to (4). Using the results, a new fuzzy type image is generated which did not include an exact edges or non-edges of the image dataset. Therefore, we applied median filter (size 3×3) on this novel image dataset and then image edges are highlighted or enhanced. This study, a median filter technique is utilized to eliminate the impulse noise in microscopic images and this will not produce any new pixel value in the image when marks the edges when compared to averaging filters.

5 Experimentations and Results

The presentation of the proposed algorithm was assessed on microscopic blood smears of malaria parasite images of different stages such as ring forms, schizonts, trophozoites, and gametocyte stages. These microscopy images are prepared routine laboratory conditions of the hospital, Hyderabad, India under the supervision of pathologist. In addition, some malaria sample images are collected from differ-

Algorithm-1 A New Filtering Algorithm for Edge Enhancement

Input: Microscopic malaria blood image dataset $A = (m, n)$.
Output: Malaria parasite edge enhanced data image.
Step 1: Collect microscopic malaria blood images infected with Plasmodium falciparum of different stages such as ring forms, schizont, trophozoite, and gametocyte stages.
Step 2: Select a microscopic images and converted into grayscale images. Identify and eliminate the saturation and hue related information in given image while retentive the luminance.
Step 3: Min-max normalization technique used to perform the linear transformation on grayscale image using formula discussed in Sect. 3 in Eq. (5).
Step 4: Compute the membership function using Eq. (6) on normalized image dataset using step-3.
Step 5: Calculate lower and upper membership using Eqs. (3) and (4) with $\alpha = 0.37$ from step-4.
Step 6: Measure the new membership function using Eq. (7) with the help of step-5. A novel fuzzy image is generated which is not included an exact edge or non-edge image dataset.
Step 7: Apply median filter on this new membership function from derived from **Step-6**, and then edges of the new image are highlighted or enhanced.
Step 8: Repeat the Steps (2)–Step (8) for each image in data image.
Step 9: Finally, malaria parasite edge enhanced image obtained.
Step 10: Stop.

ent websites through online. In Fig. 2(a–c) presented a sample implementation of proposed algorithm on microscopic blood image. It clearly, we observe that with $0 \leq \lambda \leq 1.87$ edges are enhanced of the microscopic images which visually represent the better results other than $\lambda = 2$.

The parameter α is usually determined heuristically that belongs to [0,1], in this experiment we chosen $\alpha = 0.37 \in [0, 1]$. Gaussian membership values may be utilized by taking the standard deviation of the membership functions, i.e., lower and upper membership function results presented in Fig. 3.

From these outcomes, clearly, we observed that the microscopic blood images are infected parasites and edges are accurately enhanced with proposed method. Even though Sobel edge detection method and rank-ordered filter method, the edges are not clearly visible.

Figure 3a is grayscale of Plasmodium falciparum with schizonte stage image. Figure 3b shows type II lower membership values in terms of ambiguity and Fig. 3c presents type II upper membership values in terms of vagueness of image blood cells. Using the membership functions, the grayscale values is converted into fuzzy values. Figure 3d shows the Sobel edge operator approach, but this is not clear visual quality of malaria parasite edges in the image. Figure 3f represents the rank-ordered filtered method, which shows that the edges are not clearly highlighted. Figure 3e shows that edge boundaries are much more conspicuous using the proposed method. In terms of the assessment of malaria diagnosis, it is also important an estimation of the life stage of malaria parasite that infected erythrocyte.

Figure 4 shows illustrations sample of Plasmodium falciparum with ring stage, which can be confused with parasites in early stages. However, ring stage is difficult to segment because of overlapping edges with chromatin dots. Figure 4b, d parasite not clearly visible with Prewitt and rank-ordered filtered methods. However, in Fig. 4c

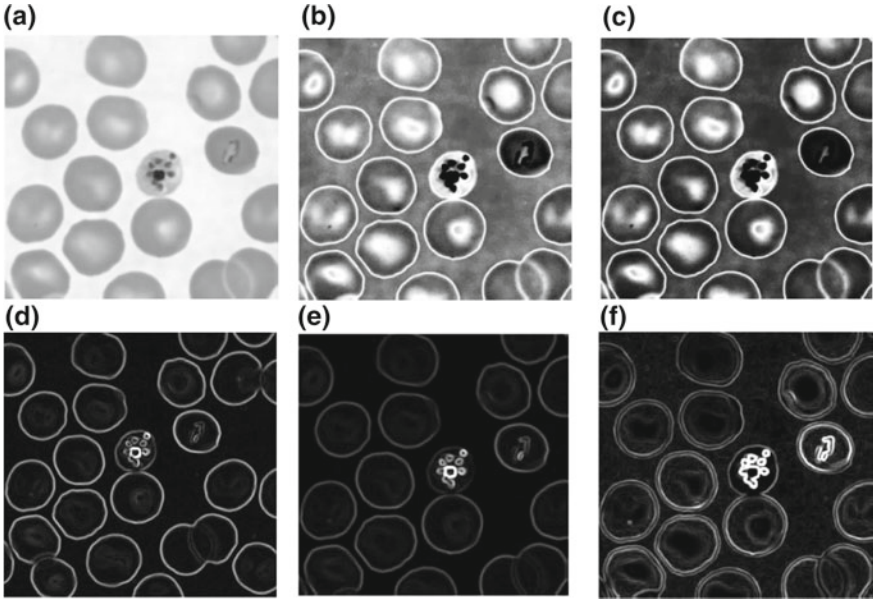


Fig. 3 **a** Original *P. falciparum* the schizonte image, **b** Gaussian lower membership function, **c** Gaussian upper membership function, **d** Sobel edge method, **e** Rank-ordered filtered method, **f** Proposed method

shows the results of Sobel filter method with clear edges only infected regions with blurry structures, which is quite better than Prewitt and rank-ordered filtered methods. Figure 4e demonstrated the outcomes of the proposed where image clearly enhanced edges that are much better than other methods with original image. Figure 5a indicates an image of *P. falciparum* of gametocyte stages, which is not clearly visible parasite.

Figure 5b illustrates the test outcomes of Prewitt filter method does not show the clear edges which shown some edge are broken also few were not clearly visible. Figure 5c illustrates the test results of using Sobel edge filter method that had slightly better than the Prewitt filter method. Figure 5d shows the test results of rank-ordered filter method, and it is observed that this approach does not enhance all the edges of the image and it enhanced on the infected gametocyte cell image. Figure 5e displays the results of proposed method, which performs well than Prewitt filter, Sobel edge filter, and rank-ordered filter methods. Finally, it concludes that the proposed approach which makes superior visual quality of edge enhancement among the existing methods.

Figure 6a shows an image of *P. falciparum* with trophozoite stage stages, which is not clearly visible parasites. Figure 6b shows the test results of Prewitt filter method which does not show the clear blood cell edges which shown some edges are broken also some cell edges are not clearly visible. Figure 6c shows the test results of using Sobel edge filter method that had slightly better edge enhancement than the Prewitt

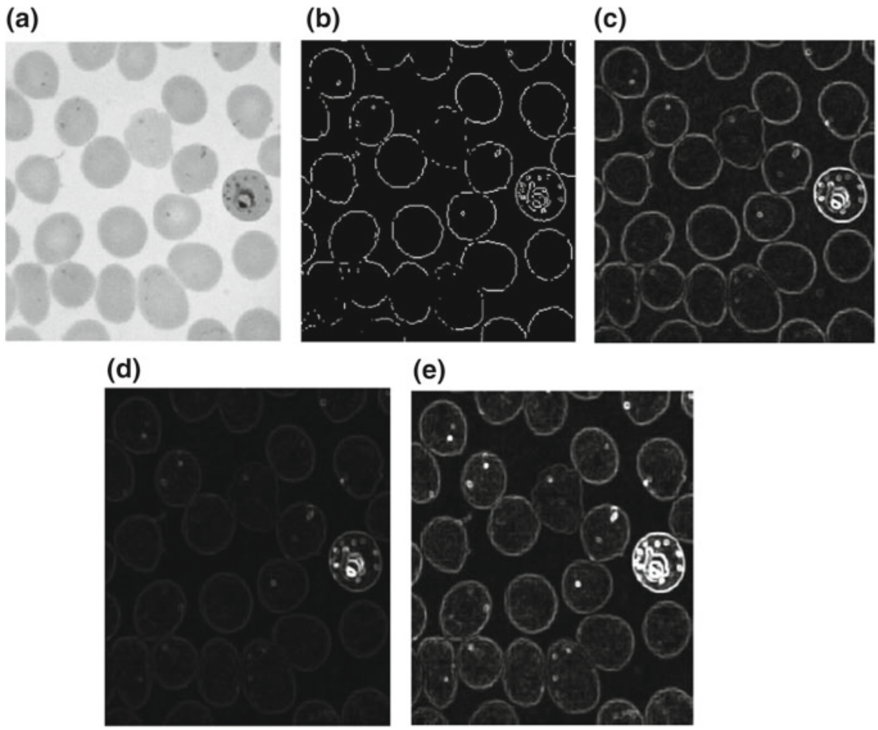


Fig. 4 **a** Represent the *P. falciparum* image with ringform stage, **b** Edge enhancement image based on Prewitt method, **c** Edge enhancement image based on Sobel edge method, **d** Edge enhancement image based on rank-ordered filter method, **e** Edge enhancement image of our method

filter method. Figure 6d shows the test results of rank-ordered filter method, it is observed that this approach also does not enhance all the edges of the blood image and it enhanced on the infected gametocyte cell image. Figure 6e represents the test results of suggested method, which performs better than Prewitt filter, Sobel edge filter, and rank-ordered filters. Finally, it determined that the suggested technique performs better visual quality on edge enhancement among all other methods.

5.1 Quantitative Performance Analysis

After visual valuation, quantitative performance analysis is also mandatory to compute the edge enhanced medical images. There are many methods to estimate the quantitative performance measure such as Shannon's entropy, Peak signal-to-noise ratio (PSNR), index of fuzziness, mean square error rate, and other methods. In this study, Shannon's entropy and PSNR is used to estimate the quality of the blood image. The Shannon's entropy is presented as follows:

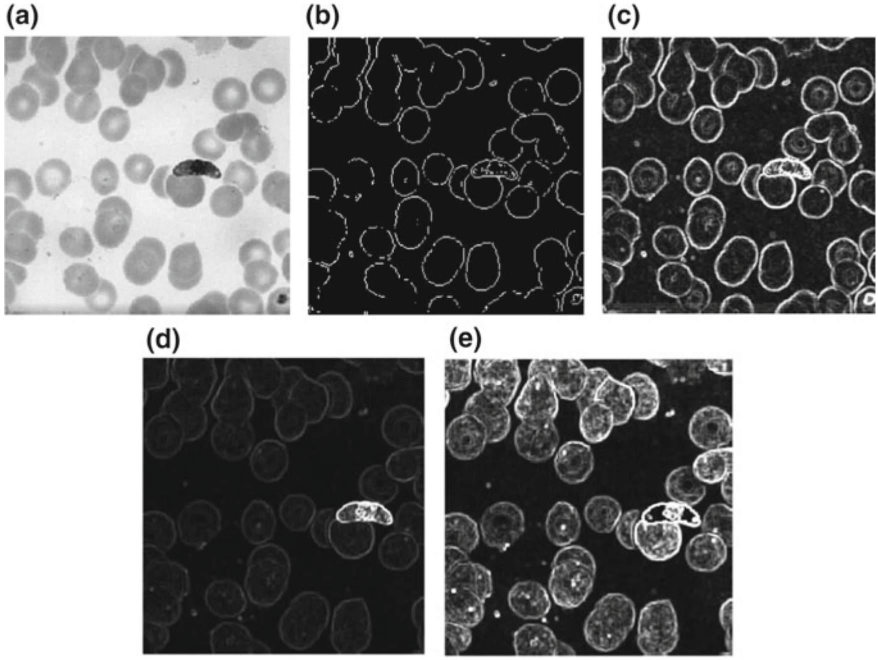


Fig. 5 **a** Represents the original *P. falciparum* with gametocyte stage, **b** Edge enhancement image based on Prewitt method, **c** Edge enhancement image based on Sobel edge method, **d** Edge enhancement image based on rank-ordered filter method, **e** Edge enhancement image based on proposed method

$$\text{Entropy}(X) = \frac{1}{(M \times N)} \sum_i \sum_j \mu_{\tilde{A}}(x_{ij}) \times \log\left[\mu_{\tilde{A}}(x_{ij})\right] - \{1 - (\mu_{\tilde{A}}(x_{ij})) \times \log(\mu_{\tilde{A}}(x_{ij}))\} \tag{8}$$

where $M \times N$ represents the size of the image and $\mu_{\tilde{A}}\left(x_{ij}\right)$ represent the membership function values of (i, j) the pixel values. PSNR is one of the standard quality measures in image analysis. In this experimentation, the PSNR value for grayscale image is defined by

$$\text{PSNR} = 20 * \log_{10}(A) - 10 * \log_{10}\left(A - \widehat{A}\right)/255 \tag{9}$$

where $\widehat{A}$ represent the edge enhanced image and A represents the original grayscale image.

Table 1 represents the entropy of microscopic blood image of malaria with different stages of malaria parasite. From this table, it is clearly observed that Prewitt filter shows the average entropy of 0.783425, Sobel edge filter demonstrate average entropy of 0.868975, and rank-ordered filter shows average entropy of 0.8081,

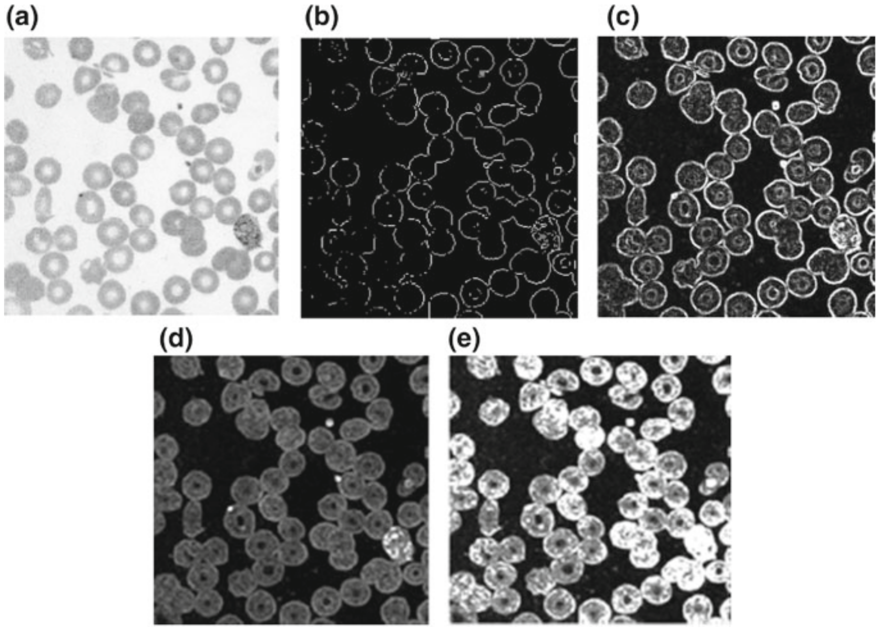


Fig. 6 **a** Represent the original *P. falciparum* with Trophozoite stage, **b** Edge image based on Prewitt method, **c** Edge image based on Sobel edge method, **d** Edge image based on rank-ordered filtered method, **e** Edge image based on proposed method

Table 1 Shannon's entropy values obtained while computing the quality of the malaria image dataset for different stages

Entropy measure				
Malaria dataset with different stages				
Microscopic images	Prewitt method	Sobel method	Rank-ordered filter	Proposed method
Ring forms	0.8682	0.8870	0.9235	0.9332
Gametocyte	0.7077	0.8447	0.8796	0.8912
Shizont	0.8409	0.8709	0.7578	0.8894
Trophozoites	0.7169	0.8733	0.6715	0.8948
Average	0.783425	0.868975	0.8081	0.90215

and our approach shows the average entropy of 0.90215. However, the suggested approach is significantly improved with higher entropy values when compared with other methods.

From Table 2, it clearly says that Prewitt method shows higher PSNR value for gametocyte stage with 56.3382 and lower PSNR value for trophozoites 54.1041. Sobel edge method shows higher PSNR value for ring form stage with 55.9305 and lower PSNR value tophozoites with 49.0710, rank-ordered filter method shows higher

Table 2 PSNR values obtained while computing the quality of the malaria image dataset for different stages

PSNR measure				
Malaria dataset with different stages				
Microscopic images	Prewitt method	Sobel method	Rank-ordered filter	Proposed method
Ring forms	55.2071	55.9305	54.3999	55.9601
Gametocyte	56.3382	55.3186	59.0502	59.6952
Shizont	55.0194	53.1093	57.8757	58.5438
Trophozoites	54.1041	49.0710	54.6634	54.7890

PSNR value for gametocyte with 59.0502 with lower PSNR value 54.3999 and the proposed method shows higher PSNR value for gametocyte stage with 59.6952 and lower PSNR value for trophozoites with 54.7890. However, the proposed method has higher PSNR values when compared with other edge enhanced methods.

6 Conclusions

In this study, we have demonstrated three-stage edge enhancement approach for microscopic blood images. In first stage, we converted microscopic blood image into grayscale image and normalized the grayscale image then computed Gaussian membership values with type II fuzzy sets. Second stage computes the lower membership and upper membership values of grayscale image. Using these membership values calculated a new membership function with Hamacher t-conorm on microscopic parasite images. Final stage, median filter applied on these images to obtain edge enhanced microscopic malaria parasite images. Quantitatively and qualitatively, results compared with other edge enhancement methods. It was found that the proposed approach is consistent and coherent in all microscopic malaria images with four stages, with average entropy 0.90215 and 59.69% PSNR values, respectively. It is clearly observed that the suggested technique is intelligent to determine blur and dark microscopic blood images. It can be used not only in microscopic images, but also in any other boundary edge enchantment of the image. This research is more capable for medical imaging analysis that makes sense for physicians decision for further diagnosis.

Acknowledgements The author acknowledge DRDO-DRL, Tezpur, Assam, India for providing financial support to carry out this work (Task No. DRDO-DRLT-P1-2015/Task-64).

References

1. Hassanpour H, Samadiani N, Salehi SM (2015) Using morphological transforms to enhance the contrast of medical images. Egypt J Radiol Nucl Med 46(2):481–489
2. Firoz R, Ali MS, Khan MNU, Hossain MK, Islam MK, Shahinuzzaman M (2016) Medical image enhancement using morphological transformation. J Data Anal Inf Process 4:1–12
3. Chaira T (2012) A rank ordered filter for medical image edge enhancement and detection using intuitionistic fuzzy set. Appl Soft Comput 12(4):1259–1266
4. Kumar A, Shaik F (2016) Image processing in diabetic related causes, forensic and medical bioinformatics. https://doi.org/10.1007/978-981-287-624-9_2
5. Russo F (2002) An image enhancement technique combining sharpening and noise reduction. IEEE Trans Instrum Meas 51(4):824–828
6. Kirsch RA (1971) Computer determination of the constituent structure of biological images. Comput Biomed Res 4(3):315–328
7. Rosenfeld A (1981) The max roberts operator is a Hueckel-Type edge detector. IEEE Trans Pattern Anal Mach Intell 3(1):101–103
8. Canny J (1986) A computational approach to edge detection. IEEE Trans Pattern Anal Mach Intell 8(6):679–698
9. Ulupinar F, Medioni G (1990) Refining edges detected by a LoG operator. Comput Vis Graph Image Process 51(3):275–298
10. Zhang J-Y, Chen Y, Huang X (2009) Edge detection of images based on improved Sobel operator and genetic algorithms. In: IASP 2009 international conference image analysis and signal processing, pp 31–35
11. Lei Y, Dewei Z, Xiaoyu W, Hui L, Jun Z (2011) An improved Prewitt algorithm for edge detection based on noised image. In: 4th International congress on image and signal processing (CISP), Shanghai, China. IEEE, pp 1197–1200, 15–17 Oct 2011
12. Mokhtarian F, Suomela R (1998) Robust image corner detection through curvature scale space. IEEE Trans Pattern Anal Mach Intell 20(12):1376–1381
13. Genming C, Baozong Y (1998) A new edge detector with thinning and noise resisting abilities. J Electron 6(4):314–319
14. Fürhapter S, Jesacher A, Bernet S, Ritsch-Marte M (2005) Spiral phase contrast imaging in microscopy. Opt Express 13(3):689–694
15. Chaira T (2008) A new measure on intuitionistic fuzzy set and its application to edge detection. Appl Soft Comput 8:919–927
16. Marr D, Hildreth E (1980) Theory of edge detection. Proc R Soc London B 207(1167):187–217
17. Kasturi R, Jain RC (eds) (1991) Computer vision: principles. IEEE Computer Society Press, Los Alamitos, CA
18. Santis AD, Sinisgalli C (1999) A Bayesian approach to edge detection in noisy images. IEEE Trans Circ Syst I Fundam Theory Appl 46(6):686–699
19. Kuo YH et al (1997) A new fuzzy edge detection method for image enhancement. In: 6th IEEE international conference on fuzzy systems, Barcelona, Spain. IEEE, pp 1069–1074, 1–5 Jul 1997
20. Khamy E et al (2000) A modified fuzzy Sobel edge detector. In: 17th National radio science conference, pp C32/1–C32/9
21. Lu S, Wang Z, Shen J (2003) Neuro-fuzzy synergism to the intelligent system for edge detection and enhancement. Elsevier J Pattern Recogn 36:2395–2409
22. Wu J, Yin Z, Xiong Y (2007) The fast multilevel fuzzy edge detection of blurry images. IEEE Signal Process Lett 14(5):344–347
23. Bustince H et al (2009) Interval valued fuzzy sets constructed from matrices: application to edge detection. Fuzzy Sets Syst 160:1819–1840
24. Barrenchea E et al (2011) Construction of interval valued fuzzy relation with application to generation of fuzzy edge images. IEEE Trans Fuzzy Syst 19(5):819–830
25. Chaira T (2012) A rank ordered filter for medical image edge enhancement and detection using intuitionistic fuzzy set. Appl Soft Comput 12(4):1259–1266

26. Chaira T (2015) Rank-ordered filter for edge enhancement of cellular images using interval type II fuzzy set. J Med Imaging 2(4):044005
27. Tirupal T, Mohan BC, Kumar SS (2017) Multimodal medical image fusion based on Sugeno's intuitionistic fuzzy sets. ETRI J 39(2):173–180
28. Mendel JM, Bob John RI (2002) Footprint of uncertainty and its importance to type-2 fuzzy sets. In: Proceeding of 6th IASTED international conference artificial intelligence and soft computing, Banff, Canada, pp 587–592
29. Mendel JM (2017) Type-2 fuzzy sets. In: Uncertain rule-based fuzzy systems, pp 259–306
30. Mendel JM, Bob John RI (2002) Type-2 fuzzy sets made simple. IEEE Trans Fuzzy Syst 10(2):117–127
31. Mendel JM (2001) Uncertain rule-based fuzzy logic systems. Prentice-Hall, Englewood Clif fs, NJ
32. Tizhoosh HR (2005) Image thresholding using type II fuzzy sets. Pattern Recogn 38(12):2363–2372
33. Roychowdhury S, Wang BH (1998) On generalized Hamacher families of triangular operators. Int J Approximate Reasoning 19(3–4):419–439

Implementation of Virtual Trial Room for Shopping Websites Using Image Processing

Niket Zagade, Akshay Bhondave, Raghao Asawa, Abhishek Raut and B. C. Julme

Abstract In today's world, the use of e-commerce websites has increased a lot. Online shopping gives us more information and choice about different products under various categories which are readily available. The customer just has to choose between the different products, purchase them, and the product comes to the customer's doorstep. Thus, many people like to buy many things online. Clothing is one of such categories that people can buy online. But for shopping clothes, this scenario is a little different. There is a major problem that people do not know how the clothes would actually look on them and so many people avoid buying clothes online. Sometimes the customers even send back the clothes they buy as it does not look good on them. This is why a virtual trial room has to be developed so that people do not have to wait to try on the clothes physically after it is delivered. They can try clothes virtually on the virtual trial room.

1 Introduction

In current scenario, online shopping is a big boon and people show quiet a lot interest in it. Many people buy clothes online from the e-commerce websites, but there is always a challenge while buying them. The customer who wants to buy apparel does

N. Zagade (✉) · A. Bhondave · R. Asawa · A. Raut · B. C. Julme
Department of Computer Engineering, PVG's College of Engineering and Technology, Pune, India
e-mail: niketzagade198@gmail.com

A. Bhondave
e-mail: akshaybhondave26@gmail.com

R. Asawa
e-mail: raghaosa@gmail.com

A. Raut
e-mail: abhishekraut695@gmail.com

B. C. Julme
e-mail: bcj_comp@pvgcoet.ac.in

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_65

not know how it would actually look upon them. This is a major reason why some people do not buy clothes online. Some people even buy some apparel online and then later return it as it does not look good on them.

For this reasons, a virtual dressing room will be a great useful tool for many online sellers. In this software, people would know how the clothes would look on them. Shopping websites want to make the shopping process easy for the customers. However, while shopping the clothes people always have that last doubt of how the dress would look on them and so they rather end up going to the stores physically so that they can actually try the clothes before buying it.

In our project we are developing a virtual trial room for the above problem. This project will show the customers how the clothes would look on them virtually, i.e., without actually trying them. The customer just has to select the clothes that they want to try and the software will fit it virtually over their image and show how it looks upon them.

This software would be used by the shopping websites to help improve their relationship with the customer. By using this software customers would feel satisfied before buying any clothes online.

2 Literature Survey

Previously many researches have been done on virtual trial room. This work can be classified into 2 groups: (1) 2D image based and (2) 3D model based.

For 2D image based approach there is a website named Awasaba [1] which gives Internet-based-/web-based interface for virtualized clothes over a stationary model. Srinivasan [2] implemented virtual fitting room in which human silhouette extracted then wrapping of shirt and virtual fitting is done. On the basis of body contour of customer and model, yamada [3] has proposed a model for garment image reshaping and getting more reliable fitting over actual customer body.

3D model based approaches are mainly implemented when virtual humans are used as model. It consists of deformation transfer, garment shape estimation, wrinkle estimation, etc. [4, 5] For garment shape estimation [6], CNN (Convolutional Neural Network) has been used to estimate 3D vertex displacement from template mesh.

Deformation transfer copies the deformation exhibited by a source mesh onto different target mesh. Despite these progressive methodologies, customer experiences difficulties because photo-realistic rendering of 3D garments and virtual human model is not so efficient. Additionally the processing cost of such 3D model is comparatively higher than 2D image based methods. 2D based methods have an advantage of collecting data and photo-realistic rendering are easier than 3D model based approach.

Some methodologies use cloning and dress people using website [7]. In a body and garment creation method for an Internet-based virtual fitting room by Protopsaltou [8], they built a compelling, interactive, and highly realistic virtual shop. Here customers can check how the garments would look on the human body by virtually fitting the clothes over the animated bodies. They present a straight-forward garment virtualization technique. In Made-to-Measure Technologies [9], they provide a web application to give easy and fast access to and manipulate garments to facilitate the characteristics like design, pattern derivation, and sizing of garments.

In paper by Tong [10], using Kinect sensor modeling method, human body 3D modeling is done. To get the human body model three steps are followed, first with the help of anthropometric parameters model is parameterized, using the PCL library the point cloud data is processed and matched and then realistic human body model is obtained.

Ehara and Saito [11] has used a database of marker attach t-shirt images in different poses. Zhau et al. prepared animated garment in different poses and done real-time virtual trial fitting by superimposing garment over customer using Kinect [12]. But using a Kinect is not feasible for every user in the world because they cannot carry Kinect everywhere for now and it is costly. Sometimes Kinect gives distorted result because of environment condition/light disturbance/illumination.

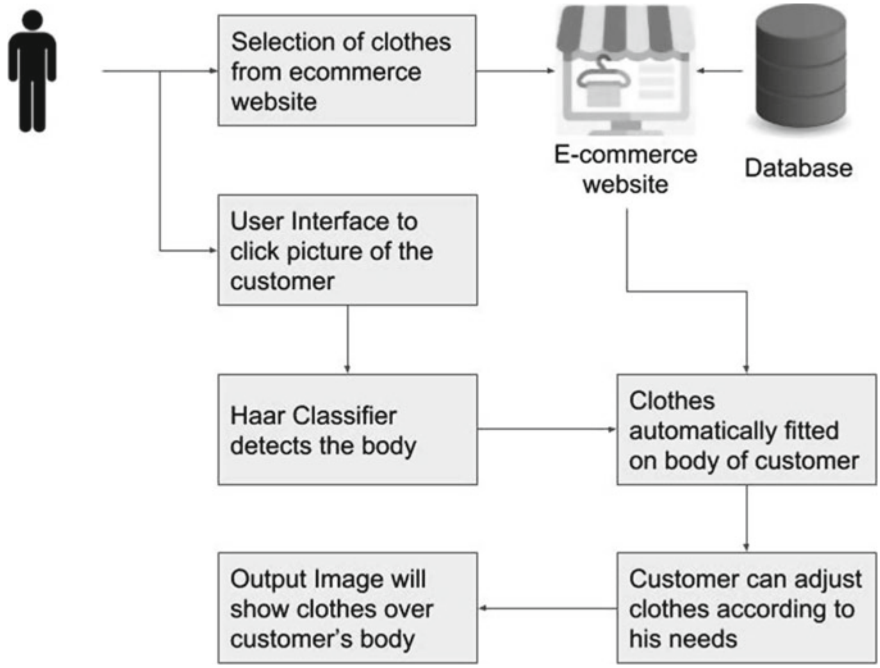
3 Proposed Work

We are developing a java application using opencv for image processing. In our project, we are implementing a virtual trial room where first we will click/take a picture of customer according to our predefined pose. The customer has to stand accordingly inside the outlined structure given in the window. We are implementing haar classifier on that image to detect full body, upper body, and lower body of customer. After that we are imposing the garments on the customer's body according to the detected parts.

The system we are developing is for the e-commerce shopping websites where the customers can select the garments they want to try. These garments will be fed to java application as input.

The software will impose garments over the image of customer's body. The use of Haar classifier is during the superimposing of garments over the body of customer. Haar classifier detects body of customer and draws a rectangular contour around body. There are three kinds in haar classifier: (1) Full body (2) Upper body (3) Lower body.

According to contour drawn by haar classifier, we will give relative position to the garments. There can be small number of false alarms in this algorithm but it is reliable for still images [13]. So we are providing a manual interface to overcome the anomalies if any occurs. The manual interface can move the superimposed garment. Customers can do it to get the proper fitting as they desire.



4 Conclusion

Several methods are used to improve the process of shopping clothes online. This software will definitely improve the customer–seller relationship and help the online sellers to expand their business. The processing speed will be better than that of Kinect and as we use webcams, the common man can easily use it. It is definitely reliable and will help both sellers and customers in their respective constraints.

References

1. Corporation, "Awaseba," www.awaseba.com/
2. Srinivasan K, Vivek S (2017) Implementation of virtual fitting room using image processing. In: IEEE International conference on computer, communication and signal processing (ICCCSP-2017)
3. Yamada H, Hirose M, Kanamori Y, Mitani J, Fukui Y (2014) Image-based virtual fitting system with garment image reshaping. In: 2014 International Conference on Cyberworlds (CW)
4. Sumner RW, Popović J (2004) Deformation transfer for triangle meshes. ACM Trans Graph (Proceeding of SIGGRAPH 2004) 23(3):399–405
5. Wang H, Hecht F, Ramamoorthi R, O'Brien J (2010) Example-based wrinkle synthesis for clothing animation. In: ACM SIGGRAPH 2010 Papers, pp 107:1–107:8

6. Guan P, Reiss L, Hirshberg DA, Weiss A, Black MJ (2012) DRAPE: DRessing Any PErson. ACM Trans Graph 31(4):35:1–35:10
7. Cordier F, Lee W, Seo H, Magnenat-Thalmann N (2001) From 2D photos of yourself to virtual try-on dress on the web. Springer, Berlin, pp 31–46
8. Protopsaltou D, Luible C, Arevalo-Poizat M, Magnenat-Thalmann N (2002) A body and garment creation method for an internet based virtual fitting room. In: Proceeding of computer graphics international 2002 (CGI '02). Springer, Berlin, pp 105–122
9. Cordier F, Seo H, Magnenat-Thalmann N (2003) Made-to-measure technologies for an online clothing store. IEEE Comput Graph Appl 23(1):38–48
10. Tong J, Zhou J, Liu L, Pan Z, Yan H (2012) Scanning 3D full human bodies using kinects. IEEE Trans Vis Comput Graph (Proceeding of IEEE Virtual Reality) 18(4):643–650
11. Ehara J, Saito H (2006) Texture overlay for virtual clothing based on PCA of silhouettes. In: Proceedings of the 5th IEEE and ACM international symposium on mixed and augmented reality, ser. ISMAR '06. IEEE Computer Society, pp 139–142
12. Zhou Z, Shu B, Zhuo S, Deng X, Tan P, Lin S (2012) Image-based clothes animation for virtual fitting. In: SIGGRAPH Asia 2012 Technical Briefs, pp 33:1–33:4
13. Kruppa H, Castrillon-Santana M, Schiele B (2003) Fast and robust face finding via local context. In: Joint IEEE international workshop on visual surveillance and performance evaluation of tracking and surveillance

Response Analysis of Eulerian Video Magnification

S. Ramya Marie and J. Anudev

Abstract The human eye has a very high optical resolution making it one of the most astonishing curiosities of the world. However, its spatial resolution is not good enough to capture everything happening around it, and it can miss out minor details, which can be termed as hidden movements. The hidden movements can be due to the extremely high speed of the visual, very small movements, long-term physical process, etc. Eulerian video magnification is a spatiotemporal video processing algorithm that can reveal hidden details that are otherwise hidden to naked eyes. In this process, a standard video sequence is spatially decomposed, and temporal filtering of the frames is done. The data so obtained as output can be used in many fields such as biomedical instrumentation, remote surveillance, etc. Here, an analysis has been done on Eulerian Video Magnification (EVM), for different video resolutions to understand its reliability.

1 Introduction

- The world around us is very dynamic and is in the process of continuous transition. The human eye is a highly complex organ which has evolved to react to a particular range of light and pressure. It can provide a three-dimensional, moving image, normally coloured in daylight. A normal human eye has a resolution of about 576 megapixels [1]. An evolutionary trait for an eye is to constantly keep collecting visual data of its surroundings and keeping track of threats, it does so by having a peripheral vision that gives it an initial impression or context before we focus on something. The peripheral vision helps us decide where to concentrate and helps us find the region of interest for our focused vision [2], such as the spotting of a

S. Ramya Marie · J. Anudev (✉)
Department of Electrical and Electronics Engineering, Amrita School of Engineering,
Amrita Vishwa Vidyapeetham, Amritapuri, Kollam, India
e-mail: anudevj@am.amrita.edu

S. Ramya Marie
e-mail: ramyamaries@am.students.amrita.edu

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_66

lurking predator. This has helped human beings in survival as a species; however, the human eye has compromised on its spatiotemporal sensitivity, and this has caused many subtle changes to go unnoticed in everyday life, such pulse rate of a person, breathing rate, sag and sway of a bridge, subtle colour changes of our skin, etc. [3]. These visuals are not really necessary for normal everyday life, but these subtle motions have great potential which can be tapped in biomedical imaging [4, 5], E-health systems [6], telemedicine [7], remote sensing, predicting of natural disasters such as avalanche, etc. The rapid improvement in the fields of biomedical imaging can be attributed to the efforts aimed at making instrumentation systems capable of sensing multiple bio-signals, non-contact type [8], accurate such as retinal imaging [9], safer for vulnerable population [10], etc.

- It is very hard to notice these small changes using a human eye; however, a computer technique can be used for visualizing such subtle changes in colour and motion variations in videos by making the variations caused by these subtle changes larger. This can be done by amplifying motions in the videos which are otherwise moved only by a hundredth of a pixel to become magnified enough to span many pixels [3]. This process is called Video Magnification. Some common video magnification algorithms are the linear Eulerian [11], phase-based [12], and Riesz [13] algorithms. This is similar to how a microscope amplifies an optical image. Using video magnification algorithm, subtle colour changes, motions, etc., can be obtained from a video that seems static. These subtle signals can be quantitatively analysed for obtaining other parameters such heart rate, respiratory rate, reconstruct sound from an object by measuring the vibrations of the object on a high-speed video, etc.
- The video magnification algorithm to be used for these purposes need to be efficient and robust. Eulerian Video Magnification (EVM) is a linear video magnification technique and is both robust and efficient. It gives better results than the motion magnification [14] where a small motion is amplified by computing per-pixel motion vector and then displacing the pixel value by magnified motion vectors. This technique yields very good results and has very high reliability; however, it is computationally expensive and any error in motion analysis would generate artefacts in the outputs which are motion magnified. Thus, this technique with high potential risk gives extremely deviant output in case of any unexpected errors.
- The most basic version of the spatiotemporal video processing has been considered here for analysing the reliability of this algorithm. It identifies the intensity variations over time for each pixel and amplifies them. The technique is used to identify subtle colour changes by measuring the colour intensity of a particular pixel.

2 Methodology

2.1 Experimental Set-up and Procedure

The experiment is aimed at studying the reliability of Eulerian video magnifications in analysis or study of subtle colour changes that are usually hidden to human eyes in daily life. Three standard optical videos are considered here. They are of 144p, 180p and 192p in resolution, respectively. These optical videos can be considered as static optical videos with no visible changes that can be detected by a naked human eye.

The static video considered here was that of a human cheek which had no visible changes that could be observed. This subject was chosen so as to illustrate the potential of EVM in the field of biomedical instrumentation. The subject of the video was made to sit still, and the video was taken for a duration of 15 s. This process was done using three cameras of different resolutions at the same time. Human cheek is subject to constant colour changes due to the perfusion of blood through the face. The human face is redder when the heart pumps blood and is more of a shade of yellow than red when the heart contracts. This is because the blood rushes through the face [15] as the heart compresses, causing the face to gain a colour shade from the colour of blood and when the heart contracts the colour of the skin is more prominent. However, this colour changes are very subtle and hence not noticeable by human eyes.

For this experiment, the subject is advised to be at rest for a duration starting from at least 15 min prior to which the video was taken. The cameras are mounted on a tripod stand; this is for preventing any error due to external vibrations. It is would be better if the frame rate of the camera is high.

2.2 Eulerian Video Magnification (EVM)

The basic idea of colour magnification in Eulerian video magnification is to amplify variation of colour values at any point or spatial location (pixel) in a temporal frequency band, which suits a certain phenomenon [16]. A common example of EVM has been illustrated below. Here, a particular band of the temporal frequencies for skin colour has been amplified. This provides the variation of redness as the blood flows through the face. These variations have been in turn used to obtain human heart rate.

The technique used here to extract and reveal the required signal is localized spatial pooling and bandpass filtering. The whole process is schematically shown in Fig. 1. To obtain a decent quality of the output video in a limited time period, before analysing, down-sample the videos and filter it using a spatial low-pass filter. This reduces the noise and to boosts the subtle changes in the video. Then this video is decomposed into different bands of wavenumbers k ($k = 2\pi\lambda$).

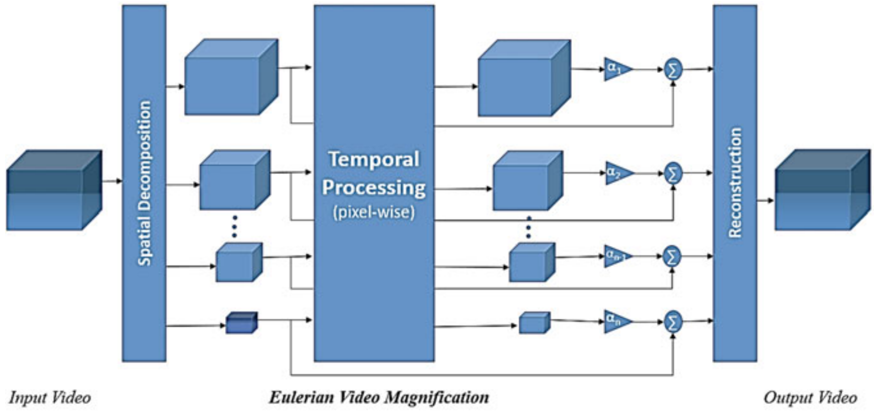


Fig. 1 Schematic diagram of EVM

Spatial processing is to be then followed by temporal processing, which is to be performed on each band of wavenumbers. Applying a bandpass temporal filter to each band of wavenumbers helps in extracting (pass) the motions that suit the frequency bands of the observed phenomenon. Magnification of motion of different passed wavenumbers is different because of two reasons. First, the signal-to-noise ratios corresponding to every wavenumber could be different. Second, some wavenumbers considered might not hold true for the linear approximation used in motion magnification. In the second case, the magnification is reduced to suppress artefacts. After the magnification is completed, these magnified bandpassed signals are added to the original signal to observe the effect of motion/colour magnification.

The process can be simply put as:

(1) The video is first decomposed into different bands of wavenumbers.
(2) The same temporal filter is then applied to all bands of wavenumbers, to reveal the time interval of the motion and the motion of each band of wavenumbers.
(3) Then a passband filter is applied to pass the bands of wavenumbers that suit the time interval of the observed phenomenon.
(4) The filtered bands of wavenumbers are then amplified by a given factor α, and then added back to the original signal and collapsed to generate the output video.

2.3 Filter and Amplification Criteria

The principle of Eulerian video magnification depends heavily on its filters and amplification factors. Procedures for selecting the filters and amplification parameters [16] are:

(1) Initially, a temporal bandpass filter that suits the phenomenon under observation was selected, to extract the desired motions or signals. The choice of temporal filter depends on the signal to be extracted; for example, temporal filters with broad passbands are used for broad but subtle motion magnification, narrow passband filters are used for colour amplification, etc.

(2) The user then selects the amplification factor α and a wavenumber cut off (specified by spatial wavelength, α) beyond which an attenuated version of α is used. α can be forced zero for all wavelength less than the spatial wavelength.

3 Simulation Results of Eulerian Video Magnification

The simulation was done using MATLAB. Here, a pre-recorded video was used. The output video from EVM was saved as a sequence of frames and was stored framewise into the hard disk as shown in Fig. 2. This output video was then further used for plotting mean light intensity of a primary colour to that of the number of frames in the video. The aforesaid video was 15 s long and had 104 frames. At first, the 144p video was considered, and a graph was plotted using data from EVM output of the video, which described the relationship between the intensity of light for a particular pixel with respect to the frame.

The graph shown in Fig. 3 draws comparisons between the output of EVM for the three different resolutions.

The output of the Eulerian video magnification for a static video consistently gave the same value as output irrespective of the number of times the video was processed over and over again. Also, for different resolutions, it has been observed that the Eulerian video magnification output tend to follow the input and didn't show much deviation from each other.

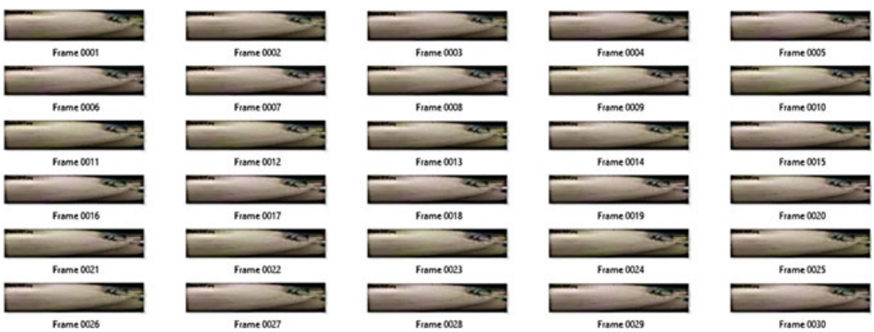


Fig. 2 Frames of the output video of Eulerian video magnification

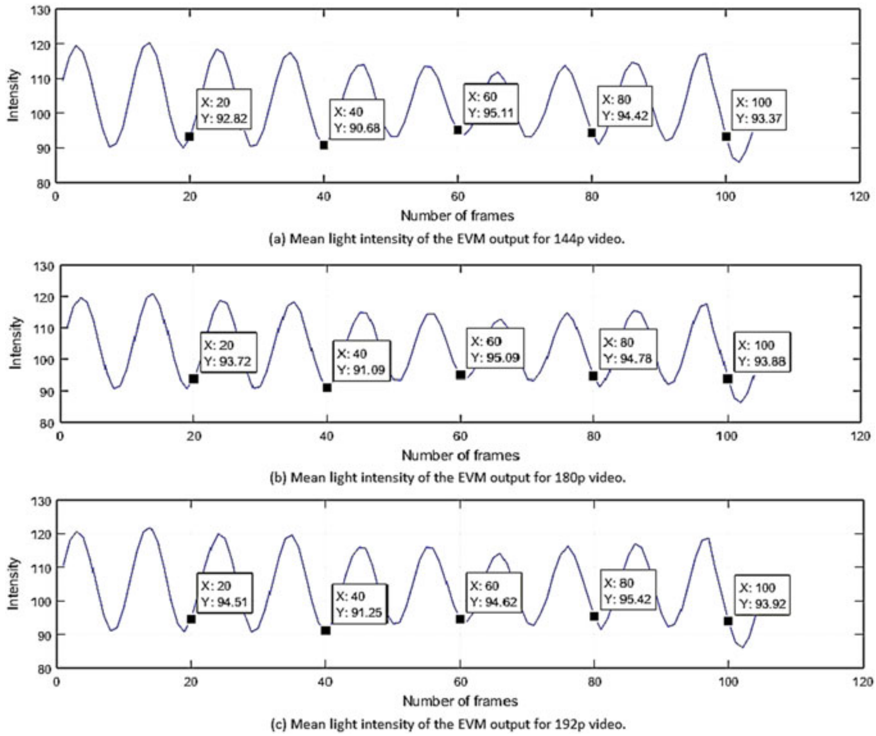


Fig. 3 Plots of mean intensity of a primary colour against the number of frames of the output of EVM for different video resolutions

4 Comparison of EVM Outputs for Different Resolutions

The light intensity of a primary colour pixel of Eulerian video magnification output for different resolutions has been plotted here. Comparisons have been drawn between light intensity at 144p and 180p in Table 1 and light intensity at 180p and 192p in Table 2. The difference between the light intensity for a particular frame for the outputs of the EVM has been compared and their difference has been found. It has been found that the magnitude of the difference between the resolutions with each other is always within the range of ± 1. This difference is marginal and won't make much difference when it comes to major instrumentation systems when calculating error percentages, and they can be reduced by proper digital filtering and can still well within engineering standards [17, 18].

Table 1 Comparison between the light intensity of a particular primary colour pixel for EVM output of 144p and 180p videos

Frame	Light intensity		Deviation (±)
	144p	180p	
20	92.82	93.72	0.9
40	90.68	91.09	0.41
60	95.11	95.09	−0.02
80	94.42	94.78	0.36
100	93.37	93.88	0.51

Table 2 Comparison between the light intensity of a particular primary colour pixel for EVM output of 180p and 192p videos

Frame	Light intensity		Deviation (±)
	180p	192p	
20	93.72	94.51	0.71
40	91.09	91.25	0.16
60	95.09	94.62	−0.47
80	94.78	95.42	0.68
100	93.88	93.92	0.04

5 Conclusion

It has been observed that the graphs showed the same values irrespective of the number of times the program has been repeated. This shows that the algorithm fulfils the major instrumentation criterion such as repeatability, reproducibility, accuracy, reliability, etc. It has great fidelity of input to output ratio, and hence irrespective of the camera, the results are bound to be proportional. Upon using thermal imaging, the problem involving external factors such as light intensity, fluctuations due to fluctuations in the inconsistent light source can be avoided. In the future, this algorithm can find applications in biomedical instrumentation systems for measuring subtle processing in the human body such as heart rate, clots, respiratory rates, etc.

Acknowledgements The authors would like to express their gratitude to Dr. Ravikumar Pandi, Project Coordinator, Assistant Professor, Dept. of Electrical and Electronics Engineering, Amrita Vishwa Vidyapeetham, Amritapuri for his continuous support and motivation, and Chairperson Dr. Manjula G. Nair, Dept. of Electrical and Electronics Engineering, Amrita Vishwa Vidyapeetham, Amritapuri for providing all opportunities and facilities for the fulfilment of this work.
The authors also express their gratitude to the panel of reviewers who helped in reviewing the work and helped in organizing the contents.

References

1. Clarkvision photography [Online]. Available: http://www.clarkvision.com/articles/eye-resolution.html
2. The complexities of the human eye from the blind spot and macula to focused and peripheral vision. [Online]. Available: https://www.zeiss.co.in/vision-care/enin/better-vision/understanding-vision/eye-and-vision/the-complexities-of-the-human-eye-from-the-blind-spot-and-macula-to-focused-and-peripheral-vision.html
3. Wadhwa N, Wu H-Y, Davis A, Rubinstein M, Shih E, Mysore GJ, Chen JG, Buyukozturk O, Guttag JV, Freeman WT et al (2016) Eulerian video magnification and analysis. Commun ACM 60(1):87–95
4. Bennett SL, GoubranR, Knoefel F (2016) Adaptive Eulerian video magnification methods to extract heart rate from thermal video, pp 1–5
5. Menon HP, Narayanankutty KA (2016) MRI/CT image fusion using gabor texture features, vol 530, pp 47–60
6. Huang A, Xie L (2015) Healthinfo engineering: technology perspectives from evidence-based mhealth study in we-care project. Int J E-Health Med Commun 6(1):22–35
7. Liu L, Stroulia E, Nikolaidis I, Miguel-Cruz A, Rincon AR (2016) Smart homes and home health monitoring technologies for older adults: a systematic review. Int J Med Inf 91:44–59
8. Nilakant KR., Menon HP, Vikram K (2017) A survey on advanced segmentation techniques for brain MRI image segmentation. Int J Adv Sci Eng Inf Technol (Insight Society) 7(4):1448–1456
9. Menon HP, Gayathri V (2017) Vasculature detection from retinal color fundus images using linear prediction residual algorithm. Int J Pure Appl Math (Academic Press) 114(12):171–178
10. Vadivelu S, Ganesan S, Murthy OR, Dhall A (2016) Thermal imaging based elderly fall detection
11. Wu H-Y, Rubinstein M, Shih E, Guttag J, Durand F, Freeman W (2012) Eulerian video magnification for revealing subtle changes in the world
12. Wadhwa N, Rubinstein M, Durand F, Freeman WT (2013) Phase-based video motion processing. ACM Trans Graph (Proceedings SIGGRAPH 2013) 32(4)
13. Wadhwa N, Rubinstein M, Durand F, Freeman WT (2014) Riesz pyramids for fast phase-based video magnification. In: 2014 IEEE international conference on computational photography (ICCP). IEEE, pp 1–10
14. Liu C, Torralba A, Freeman WT, Durand F, Adelson EH (2005) Motion magnification. ACM Trans Graphics (TOG) 24(3):519–526
15. Edwards DJ, Cattell M (1930) The action of compression on the contraction of heart muscle. Am J Physiol Legacy Content 93(1):90–96
16. Brecelj T (2013) Eulerian video magnification. In: University of Ljubljana Faculty of Mathematics and Physics, pp 1–15
17. Instrumentation error calculation and setpoint determination. Engineering Standard ES-002, pp 1–14 (1994)
18. Rajevencelta J, Kumar CS, Cattell M (2016) Improving the performance of multi-parameter patient monitors using feature mapping and decision fusion. In: Region 10 conference (TENCON). IEEE, pp 1515–1518

Face Authentication and IOT-Based Automobile Security and Driver Surveillance System

Mahesh R. Pawar and Imdad Rizvi

Abstract Automobile industry is one of the largest and fastest growing industry and the actual reason behind it is, up-growing men to vehicle ratio. A lot of new vehicles are coming in the market and people are using them by spending substantial amount of money. This increasing ratio of man to vehicle is squeezing the crimes regarding vehicle robbery and accidents even though there are lots of safety features readily available. Hence this paper proposes simple low cost solution, based on strong biometric mechanism that involves face authentication. The system that uses night vision camera to capture the face of person seating on the driver's seat and some sensors to provide his surveillance in the accidental situations. This system also gives us instant alert with latest captured image of vehicle's interior on email.

1 Introduction

Automobile industry has tremendously grown up in the past years and still it is growing. It has become one of the largest as well as fastest growing industry and there is huge potential in it. Plenty of companies every year introduces new cars with many exciting and innovative features for luxury, security and also for comfort. The automobile industry is getting upgraded with the help of automation and moving towards new era. Many companies have already introduced their automated electric cars. If we observe that, even after having these features, there are still cases of vehicle thefts and misuse of vehicles has not reduced but increased. People die due to lack of surveillance systems, Many criminals walk free because of lack of proper evidence against them. Government, transport authority. The police are trying their best to overcome these crimes, Hence it is a time to upgrade the vehicle by itself with

M. R. Pawar (✉) · I. Rizvi
Department of Electronics and Telecommunication Engineering,
Terna Engineering College, Navi Mumbai, India
e-mail: maheshrpawar9@gmail.com

I. Rizvi
Electrical Engineering Division, Higher Colleges of Technology, Sharjah, UAE

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_67

simple, low cost, easily available and trustworthy mechanism, which will actually help to give guaranteed security and surveillance of vehicle.

2 Literature Review

2.1 Tracking Method

If we consider last few years of work on automobile security, many techniques and methods are evolved. One among them is GPS (Global Positioning System) [1, 2] sensor which gives position of automobile in altitude and longitude. Then this data is transferred over air to owner or cop. This wireless transmission is mostly done with IP based modules like GPRS (General Packet Radio Service) to make IoT-based system or using the GSM (Global System for Mobile Communication) module [3, 4]. Figure 1 can give brief idea about it [5, 6].

2.2 Biometric Authentication

The biometric authentication-based methods are like, use of finger scanner, retina scanner, voice recognition, authentication, and giving commands [7–9], then one of the most acceptable method, that is face authentication. With above biometric methods they also implement security alarm or cutting off fuel-supply, locking doors, sending images to other side or owner. All transmission of data is mostly made by the SMS (Short Message Service) or MMS (Multimedia Messaging service) through GSM or GPRS modules [10–16].

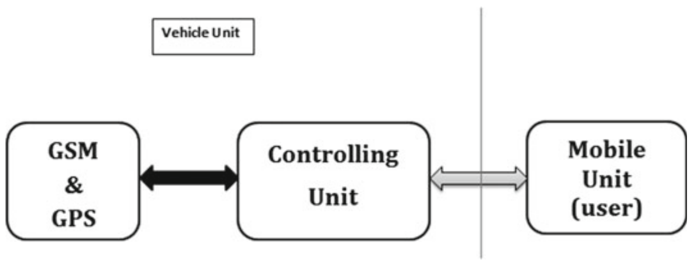


Fig. 1 Simplified block diagram GSM and GPS based system

3 Proposed Solution

Low cost with high reliability is primary objective of proposed system. The solution is conceptually much simple. Authors intention is to provide low cost, less complex, highly reliable system and At the same time it must be user friendly so that to a lay man can handle it. The person trying to access the vehicle must get identified first and then should get authorized. The following embedded system block diagram in Fig. 2 is the proposed idea with increased simplicity and functionality.

The proposed system requires 5 V/2 A power supply. Camera is used which acts like a transducer and it will take image and will provide suitable form of image in electric signal. This data will be send to processor for further processing and operation. We have used 8MP night vision Noir camera having maximum resolution of "3280 × 2464". Still we are taking pictures with resolution of "1024 × 768". Sensor detects the vibrations and makes output low or high, as it directly connected by wire to the controlling unit as it is shown in Fig. 2.

Controlling unit plays vital role, which takes all inputs and also process it. Raspberry pi 3b is used as development board. Vehicle control is an internal part of vehicle which include the key component of the vehicle, that is an ignition system. It will be handled by the controlling unit. Owner's device means it could be the mobile or Computer from where user can access an email account.

Raspbian stretch Operating System through noobs as per standard procedure is used. Python language is used for programming and VI editor for writing as well as editing code are used. OpenCV software is considered for the image processing. For face Recognition there are 3 algorithm-based functions made available in OpenCV. These are "Eigenfaces" second one "Fisherfaces" and third one is "Local Binary Patterns Histograms" (LBPH).

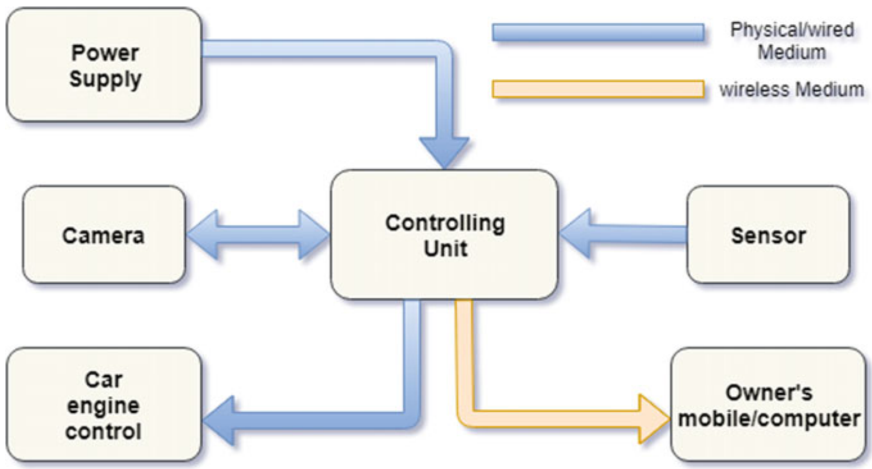


Fig. 2 Simplified block diagram of system

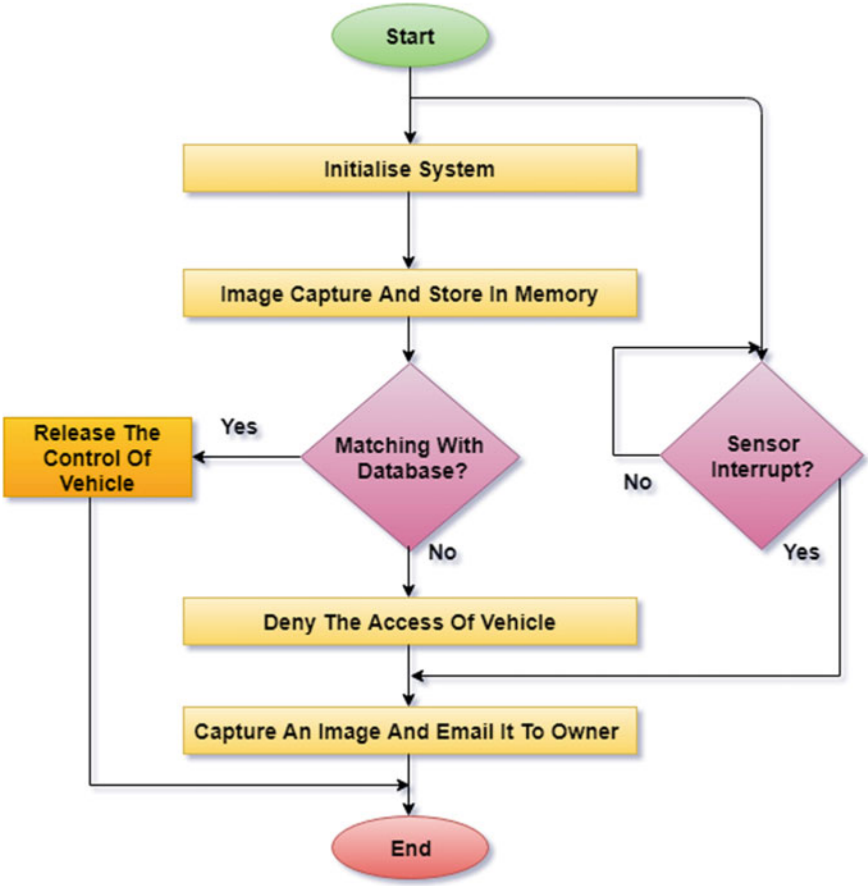


Fig. 3 Flowchart of proposed system

"Fisherfaces" and "Eigenface's" finds a mathematical relation and description of the most prior dominant property of our training image database as a whole. LBPH analyzes every face in training image database individually and independently.

The driver's image is captured via camera. Controller will do the comparison of recent image captured by the camera and the pre-stored images of authorized person as shown in Fig. 3. If the image matches, then controller will release the access of vehicle for authorized person, else controller denies the access of vehicle and it will send the image of unidentified person through an email to an authorized person. The vibration sensor will be continuously monitored for the vibrations. If any sudden vibrations are detected at any instant then camera will capture the image and send it to owner, on his email ID.

4 Result Analysis

Proposed system uses LBPH face recognition as it is simple and gives better result in different angle of views conditions. Proposed system also uses night vision camera which results in excellent face matching results in darkness too.

Figure 4 shows the database of vehicle owner. There could be multiple authorized individuals as per owners wish. Figure 4 indicates database of two authorized individuals.

Figure 5 shows face detection as well as authentication. Proposed system has successfully identified the authorized persons with their names.

Figure 6 shows that the proposed system has successfully identified an unknown person whose image was not in the database. Then it sends an email to an owner containing alert message with the image of the unauthorised person.

Similarly, the camera takes the picture after sensing heavy vibrations by vibration sensor. As soon as the proposed system senses the vibrations instantly sends the alert

Fig. 4 Database of Mahesh and Utkarsh

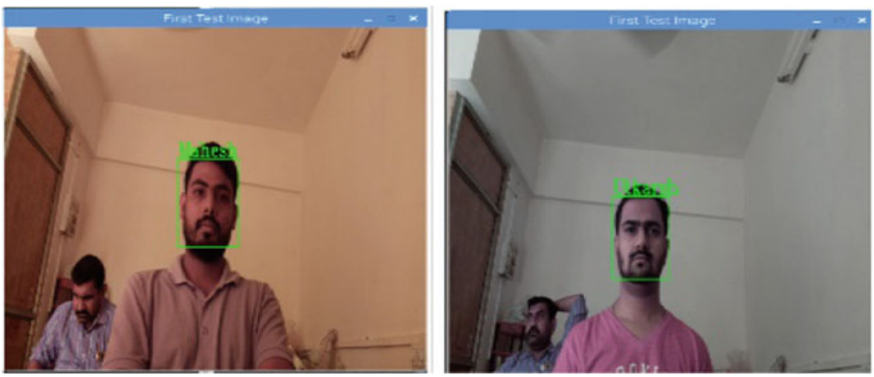


Fig. 5 Known person detected

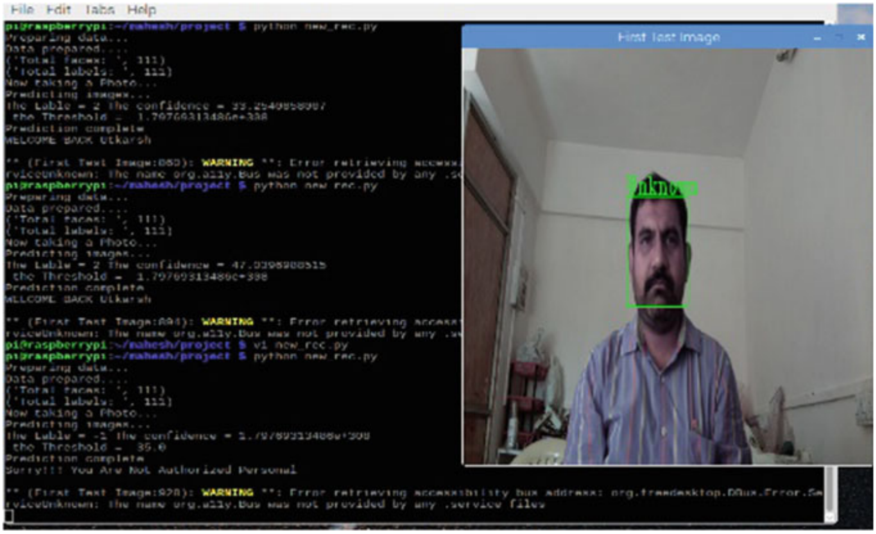


Fig. 6 Unknown person detected

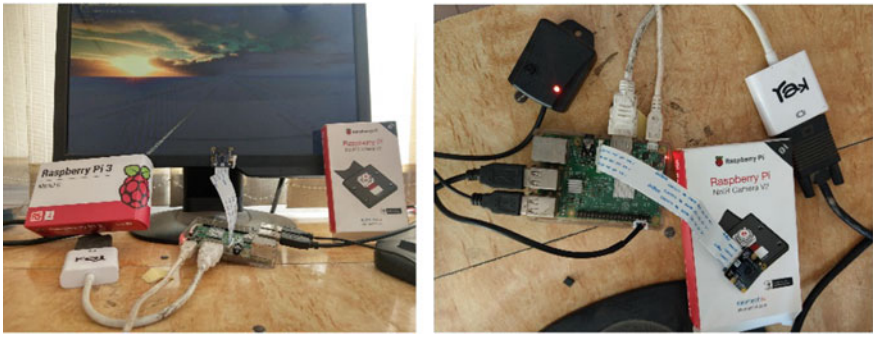


Fig. 7 Complete hardware setup

message with picture of the interior of the vehicle. The actual setup containing all the hardware is shown in Fig. 7.

5 Conclusion

The proposed system easily distinguishes known and unknown person. The use of night vision camera gives excellent results even at the night time. Developed system completes the procedure within few seconds. The owner of car can easily park his vehicle anywhere as he gets instant security alert even for vibrations in vehicle.

The proposed system gives alert at low cost as well as evidence of unauthorised person. It also provides driver surveillance. This system works at very low power like on portable power bank used for the mobile charging. Hence it works like a silent watchman for an automobile.

Acknowledgements The authors would like to take the opportunity to thank all those individuals who have contributed or helped in some way in the preparation of this text. The authors would like to take the opportunity to thank all those individuals who have contributed or helped in some way in the preparation of this text. Special thanks to Mr. Vijay Kamble for their guidance, technical support and for providing me platform. I thank to Utkarsh chatte and again to Mr. Vijay Kamble for permission to use their faces for this research.
In my department, I thank to Dr. Imdad Rizvi for their academic support. For grammar and language support I thank to Prof. Geetanjali yampalle. I would also like to thank to Mandakini Chede and Apurav Patil for encouraging me.

References

1. Liu Z, Zhang A, Li S (2013) Vehicle anti-theft tracking system based on Internet of things. In: IEEE international conference on vehicular_electronics and safety (ICVES), Dongguan, pp 48–52. https://doi.org/10.1109/icves.2013.6619601
2. Shruthi K, Ramaprasad P, Ray R, Naik MA, Pansari S (2015) Design of anti-theft vehicle tracking system with a smartphone application. In: 2015 international conference on information processing (ICIP), Pune, pp 755–760. https://doi.org/10.1109/infop.2015.7489483
3. Lita I, Cioc IB, Visan DA (2006) A new approach of automobile localization system using GPS and GSM/GPRS transmission. In: 29th international spring seminar on electronics technology, St. Marienthal, pp 115–119. https://doi.org/10.1109/isse.2006.365369
4. Bavya R, Mohanamurali R (2014) Next generation auto theft prevention and tracking system for land vehicles. In: 2014 international conference on information communication and embedded systems (ICICES), Chennai, pp 1–5. https://doi.org/10.1109/icices.2014.7033987
5. Hameed SA, Khalifa O, Ershad M, Zahudi F, Sheyaa B, Asender W (2010) Car monitoring, alerting and tracking model: enhancement with mobility and database facilities. In: 2010 international conference on computer and communication engineering (ICCCE), Kuala Lumpur, pp 1–5. https://doi.org/10.1109/iccce.2010.5556796
6. Ajaz S, Asim M, Ozair M, Ahmed M, Siddiqui M, Mushtaq Z (2005) Autonomous vehicle monitoring & tracking system. In: 2005 student conference on engineering sciences and technology, SCONEST 2005, Karachi. https://doi.org/10.1109/sconest.2005.4382882
7. Bagavathy P, Dhaya R, Devakumar T (2011) Real time car theft decline system using ARM processor. In: 3rd international conference on advances in recent technologies in communication and computing (ARTCom 2011), Bangalore, pp 101–105. https://doi.org/10.1049/ic.2011.0059
8. Muji SZM, Wahab MHA, Zin MAM, Ayob J (2008) Simulation of smart card interface with PIC for vehicle security system. In: International conference on computer and communication engineering 2008, ICCCE 2008, pp 878–882
9. Ahilan A, James EAK (2011) Design and implementation of real time car theft detection in FPGA. In: 2011 third international conference on advanced computing, Chennai, pp 353–358. https://doi.org/10.1109/icoac.2011.6165201
10. Padmapriya S, KalaJames EA (2012) Real time smart car lock security system using face detection and recognition. In: 2012 international conference on computer communication and informatics (ICCCI), Coimbatore, pp 1–6. https://doi.org/10.1109/iccci.2012.6158802

11. Sreedevi P, Nair BSS (2011) Image processing based real time vehicle theft detection and prevention system. In: 2011 international conference on process automation, control and computing (PACC), Coimbatore, pp 1–6. https://doi.org/10.1109/pacc.2011.5979056
12. Sasikumar S, Ganesan R (2014) Facial and bio-signal fusion based driver alertness system using dynamic bayesian network. In: 2014 international conference on green computing communication and electrical engineering (ICGCCEE), Coimbatore, pp 1–5. https://doi.org/10.1109/icgccee.2014.6922268s
13. Sasikumar S, Ganesan R (2014) Facial and bio-signal fusion based driver alertness system using dynamic bayesian network. In: International conference on green computing communication and electrical engineering (ICGCCEE), Coimbatore, pp 1–5. https://doi.org/10.1109/icgccee.2014.6922268
14. Rizvi I, Chawda D (2009) Simulation of antilock braking system. In: International conference on emerging trends in software & networking technologies (ETSNT'09), Noida, India
15. Kolli A, Fasih F, Machot A, Kyamakyac K (2011) Non-intrusive car driver's emotion recognition using thermal camera. In: Proceedings of the joint INDS'11 & ISTET'11, Klagenfurt, pp 1–5. https://doi.org/10.1109/inds.2011.6024802
16. Saifullah A, Khawaja H, Arsalan, Maryam, Anum (2010) Keyless car entry through face recognition using FPGA. In: 2010 international conference on future information technology and management engineering (FITME), Changzhou, pp 224–227. https://doi.org/10.1109/fitme.2010.5654862

Highly Repeatable Feature Point Detection in Images Using Laplacian Graph Centrality

P. N. Pournami and V. K. Govindan

Abstract Image registration is an indispensible task required in many image processing applications, which geometrically aligns multiple images of a scene, with differences caused due to time, viewpoint or by heterogeneous sensors. Feature-based registration algorithms are more robust to handle complex geometrical and intensity distortions when compared to area-based techniques. A set of appropriate geometrically invariant features forms the cornerstone for a feature-based registration framework. Feature point or interest point detectors extract salient structures such as points, lines, curves, regions, edges, or objects from the images. A novel interest point detector is presented in this paper. This algorithm computes interest points in a grayscale image by utilizing a graph centrality measure derived from a local image network. This approach exhibits superior repeatability in images where large photometric and geometric variations are present. The practical utility of this highly repeatable feature detector is evident from the simulation results.

1 Introduction

The problem of image registration (IR) involves techniques for matching or alignment of several images of an object or scene taken by different sensing devices in different orientations, at different scales and times. IR has variety of applications in computer vision problems such as change detection, motion analysis, image matching, object recognition, and image fusion. The availability of diverse categories of data from many application areas has led to numerous research attempts during the recent past decades. For any image registration system, the following necessary aspects must be defined: feature space, transformation model, similarity metric, and an optimization method.

P. N. Pournami (✉) · V. K. Govindan
National Insitute of Technology Calicut, Calicut, Kerala, India
e-mail: pournamipn@nitc.ac.in

V. K. Govindan
e-mail: govi.kunnumal@gmail.com

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_68

The feature space represents the information in the images before carrying out the matching process. The classes of transformations that are used to align the input images form the transformation model or search space. The choice of the transformation model from the search space is determined by the search strategy employed in the optimization process. The best choice is determined based on the similarity computed using the chosen similarity metric. The search process for obtaining the appropriate transformation parameters continues until a best acceptable match is determined between the reference and the float image [1]. The cases when structural information in images is predominant or significant, image representations using features in a concise form leading to compressed representation of data. The consequent reduction in search space and search time is the major advantage. Such representations allow faster registration of complex and even distorted images. These methods are faster when compared to pixel-based approaches. The choice of robust discriminative features, their extraction, and robust feature matching techniques are the major factors determining the success of this type of registration approaches. Feature points can either be automatically extracted or be manually selected. The robust features are invariant under rotation, translation, and scales changes. Use of discriminative features provides accurate and faster matching between float and reference images.

2 Feature-Based Image Registration

Feature-based registration algorithms are more robust to handle complex geometrical and intensity distortions when compared to area-based techniques. They can significantly reduce the execution time as less number of features is used for calculating the mapping between the input images. Feature point-based registrations effectively reflect the image structure information and they provide clear solutions to image registration. A set of appropriate geometrically invariant features forms the cornerstone for a feature-based registration framework. Because of these facts, the development of powerful feature-based registration methods has been the topic of current investigation [1].

Feature point detectors, descriptors and techniques for matching the derived features are the building blocks of the aforesaid algorithms. Feature-point or interest point detectors extract salient structures such as points, lines, curves, regions, edges or objects from the images. The mapping between the reference and the oat images is computed using the translation parameters, rotation angles, and the scaling parameters. This requires an acceptable number of matching control points from the input images, which in turn impose certain properties known as affine invariance on the extracted interest points. Such a detector will definitely improve the efficiency of a feature-based registration system when associated with a stable point descriptor. In recent years, many robust algorithms have been formulated in the category of feature-based approaches of image registration. Being the backbone of such techniques, various feature detectors have also been proposed by researchers. The following section discusses about feature point detectors.

2.1 Feature or Interest Point Detector

Feature detectors are algorithms for identifying features or points of interest in images. Points of interest are feature points that are to be detected robustly. They should be computed easily. The immediate neighborhood of a feature point has sufficient information content so that a meaningful point descriptor can be attached. Feature points should be stable under large photometric and geometric variations in the image. Feature based systems represent the current active direction of research in image registration technologies, which is the final issue investigated in this thesis. Image registration methods based on interest points or features achieve fast and high image registration.

2.2 Review of Recent Works

Zitova and Flusser [1] directed their efforts toward classifying image registration techniques according to the essential ideas, which reveals that an image registration process has the following components—detection and extraction of features, feature matching, transform model estimation, finally resampling and transformation of images. Each of these steps puts significant complications in image registration, starting from the selection of appropriate features for the given task. The features should be spread over the images and neighborhoods of feature points are rich in information content. The features usually are invariant to geometrical variations in the images and to any image degradation such as noise, blur, illumination changes, etc. Once, we get matching features in the input images, the correspondence can be established between them. This helps to derive a possible transformation, which aligns the float image with the reference image.

Harris Corner Detector, SUSAN detector, scale invariant feature transform (SIFT) detector, etc., are the major first-generation interest point detectors. The basic formulation of these detectors does not accommodate affine invariance. Later, many scientists tried to incorporate modifications to these detectors so that they become invariant to translation, rotation, scaling, camera viewpoint changes, illumination changes, noise, etc. Feature point detectors not affected by affine transformation are termed affine invariant detectors. These points are repeatable even after large distortions introduced by affine transformation [2–5]. This, in turn, motivated people to solve many computer vision applications that need feature-based matching such as object recognition, image registration, image retrieval from large databases, symmetry detection, texture recognition, etc.

One major work in this class is by Mikolajczyk and Schmid [2] where they proposed detector for interest points that are invariant under affine transformation. The algorithm is insensitive to translation, scale and shape of the immediate point neighborhood. The authors report that the proposed approach for interest point detection permits region matching even with large changes in viewpoints. Miao and Jiang [6] extracted interest points out of a nonlinear rank-order Laplacian of Gaussian filter. This algorithm can be used to extract regional structures in images where most pixels in the regions are darker or brighter than their corresponding surroundings. This permits detection of abrupt variations in illumination and geometric changes. Authors state that the approach is capable of detecting interest points efficiently with respect to repeatability and discriminative nature. The detected points of interest were applied to face recognition problem on five standard databases and depicted high recognition rates.

Criado et al. reported a graph centrality based feature point detector in [7]. Global and local centrality measures are employed. This method constructs a complex network out of the image regions created with morphological watershed operations. Though the performance of this technique is comparable with that of standard Harris and Stephens's detector, the authors report that the capability is low at over-segmented regions of images. Post-processing techniques can definitely improve the scope so that interest points are detected more sharply. A wavelet-based interest point detector for visible points is proposed in [8]. Annular color histogram and texture histogram were used to establish the affine invariance of this detector. This algorithm exhibited high precision value for an image retrieval application over a data set of 1000 images.

In 2008, Saydam et al. [9] narrated a point of interest detector with remarkable feature localization making use of non-sub-sampled Contourlet Transform. Both the local and global detectors outperformed many popular algorithms in the event of scale, viewpoint and rotational changes. This feature detector is suitable for computer vision and object recognition applications. Gevrekci and Gunturk [10] devised a feature point detector, capable of handling large photometric changes in the input images. By using the interest point detector along with a contrast-stretching operator, termed as contrast invariant feature transform, this algorithm enhances the Harris corner detector. Huge computational complexity prevents this technique to be used in practical applications.

Maver [11] developed a feature point locating technique utilizing the saliency of various regions in the image, computed at different scales. The authors reported good performances in image matching and object recognition. But this algorithm failed in many cases because it is not affine adaptable with respect to local regions. Lee and Chen [12] exploited low-level histograms computed from the images for finding interest points. Since this technique is invariant to blur and illumination changes, it could be effectively applied to image matching and action classification problems. A comparative study of four feature point detectors was performed by Martins and Carvalho [13]. In noisy images, the proposed detector showed stability when applied for image matching.

In 2010, Xie et al. [14] designed a feature point detector with scale invariance property. Harris detector is applied to detect feature points in low scale-space and these points are filtered using Hessian determinant to locate points in higher scale-space. This technique exhibited real-time response in large-scale data. Li et al. [15] presented a ranking scheme for interest point detection to identify stable local interest points. Repeatability of the interest points extracted is employed in this method. By limiting the number of stable interest points, this algorithm gives a noticeable performance in image retrieval problem.

Zukal et al. [16] reported a corroborative study of KLT detector, FAST, and Harris-Laplace detector in terms of information content. There are interest point detection algorithms which were designed specifically for image registration applications. Wen and Sheng [17] presented a framework for feature-based image registration under noisy imaging conditions, using local SIFT operator and cross-correlation information. A robust image registration algorithm combining SURF detector and FREAK descriptor is reported by Yanhai et al. in 2015 [18]. Ma et al. [19] suggested an image registration technique for synthetic aperture radar (SAR) images using modified SIFT. A registration technique for super-resolution image creation is reported by Nasir et al. [20].

The literature survey carried out opens a scope for devising an interest point detector based on local image properties. Local features strengthen image registration algorithms by handling image deformities such as rotation, scale changes, illumination changes, etc. Graphs give a very good representation of any local detail in an image. In this chapter, we present an algorithm to detect highly repeatable feature points from an image using a graph centrality measure. Utilizing graph theory, a novel interest point detector having higher repeatability, is proposed, which can be employed for extracting stable feature points to assist an image registration system.

3 Proposed Algorithm

Graph centrality measures the importance of nodes or vertices in a graph. For weighted graphs, a number of centrality measures are developed such as degree centrality, betweenness centrality, eigenvalue centrality and coseness centrality, etc. A highly efficient centrality measure known as Laplacian Centrality is presented in [21]. This metric was efficiently applied in social network analysis [22]. Image processing has also been benefitted abundantly by adopting theoretical concepts from graph theory. Graphs provide a unified representation for an image to study the global and local structural properties in detail. This, in turn, attracted researchers to devise procedures to solve many computer vision tasks utilizing graph theoretical concepts.

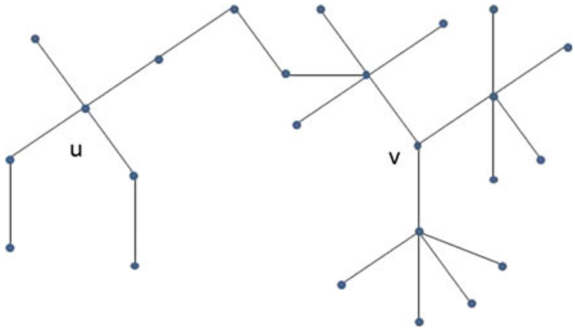


Fig. 1 A sample network where *u* and *v* are two random nodes

Consider a weighted, undirected graph or network $G=(V;E)$, where V is the finite set of nodes and E is the finite set of edges. Let $d_G(v)$ and $N(v)$ denote the degree of a node v and the finite set of nodes in the neighborhood of v, respectively. Laplacian centrality LC for the node v in a graph/network G can be calculated [22] as in Eq. (1).

$$LC(v) = d_G^2(v) + d_G(v) + 2 * \sum_{v_i \in N(v)} d_G(v_i) \tag{1}$$

In the sample network shown in Fig. 1, let us mark two nodes u and v. Even though the node u has more number of neighbor nodes than v, the Laplacian centrality makes v a more important node (central node) than u. The Laplacian centrality values $LC(u)$ and $LC(v)$ for nodes u and v are 36 and 42, respectively. This is primarily because the immediate neighbors of v have more connectivity than those of u.

Pournami and Govindan [23] explains the process of identifying the central most nodes in the weighted graphs created out of local image patches. Laplacian centrality is an intelligible measure, which can be computed in linear time. Equation 1 provides an easier way of calculating Laplacian centrality from the adjacency matrix of the network [22].

Now the validness and robustness of this centrality are to be measured in cases where the images undergo various geometric transformations, using a standard dataset. This step is important because, to be incorporated into an image registration system, the feature points detected should be stable under all image distortions. Repeatability [24] is a measure of the geometric stability of interest points among images of the same object (or scene) captured under different imaging conditions. Let I_1 and I_2 be two images of the same scene. Also let N(I_1) and N(I_2) be the number of interest points extracted from the two images I_1 and I_2, respectively. Then repeatability, R, is defined as in Eq. (2).

$$R = \frac{N(I_1) \cap N(I_2)}{\min(N(I_1), N(I_2))} \tag{2}$$

Fig. 2 Sample images from Oxford dataset

4 Experimental Results and Discussion

To demonstrate the robustness of this centrality measure, extensive experiments were designed with the algorithm written in MATLAB R2015b. The simulations were run on an Intel Core i5 (1.3 GHz) processor with 4 GB RAM and the operating system was 64-bit Mac OSX Yosemite. The robustness of this centrality measure is verified using Oxford dataset [25], where the images vary by angles of rotation, zoom or viewpoint changes, etc. Figure 2 shows sample input images from Oxford dataset.

As the first step, interest points are detected on the input images. Now, to verify the repeatability of the detected points, correspondence was established between matching points in the input images. Figures 3 and 4 show the sample output on Oxford dataset images.

Extensive experiments were conducted on more images from the Oxford dataset. The input images vary by angle of rotation, viewpoint changes, translation, illumination changes, blurring levels, etc. The repeatability of the proposed detector is calculated in each of these cases and reported in Table 1. Repeatability for speeded up robust features (SURF) points for these images is also added for comparison.

Fig. 3 Feature point correspondances on images from Oxford dataset

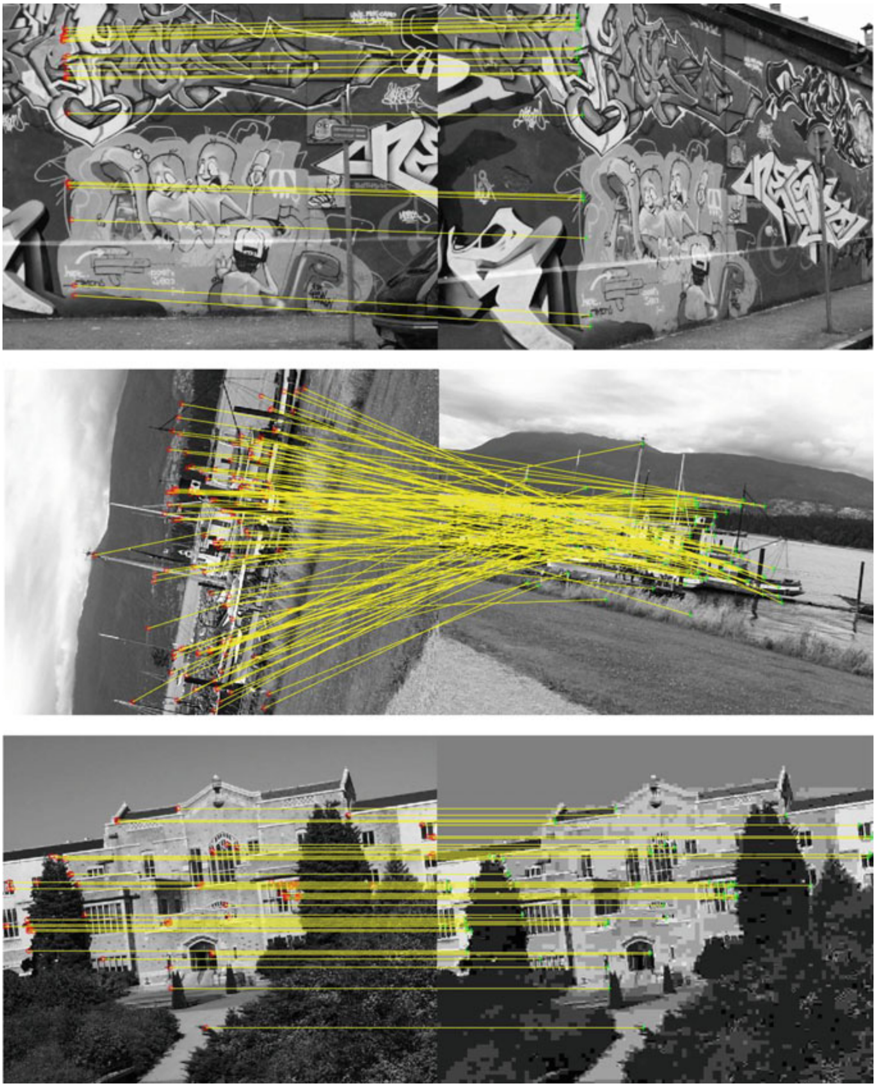


Fig. 4 Feature point correspondances on images from Oxford dataset

Table 1 Repeatability

Dataset	Nature of transformation	Proposed method	SURF
Cars (6 images)	Illumination changes	**97.98**	72.12
Boat (6 images)	Zoom, rotation	**91.76**	44.24
Building (6 images)	JPEG compression	**100**	89.98
Bikes (6 images)	Blurring	**84.26**	48.56
Graffiti (6 images)	Viewpoint changes	**96.26**	6.15
VanGogh (16 images)	Rotation	**99.29**	77.13

Repeatability values are higher for the proposed detector

5 Conclusion

The research aim of the study presented here was to test the repeatability of an efficient feature detector designed based on image geometry. We investigated the validness and robustness of this centrality measure under various geometric and photometric transformations, using standard datasets. The results obtained are highly reliable, which provide substantial evidence of the application of this centrality measure's utility computer vision applications such as image registration, image retrieval, object detection, etc. Parallel computation of node centrality values makes this approach faster.

References

1. Zitova B, Flusser J (2003) Image registration methods: a survey. Image Vis Comput 21(11):977–1000
2. Mikolajczyk K, Schmid C (2002) An affine invariant interest point detector. In: Proceedings of the 7th European conference on computer vision-part I, ECCV'02. Springer, London, pp 128–142. ISBN 3-540-43745-2
3. Lowe DG (1999) Object recognition from local scale-invariant features. In: Proceedings of the international conference on computer vision (ICCV'99), vol 2. IEEE Computer Society, USA, p 1150. ISBN 0-7695-0164-8
4. Lowe DG (2004) Distinctive image features from scale-invariant key points. Int J Comput Vis 60(2):91–110. ISSN 0920-5691
5. Mikolajczyk K, Schmid C (2004) Scale and affine invariant interest point detectors. Int J Comput Vis 60(1):63–86. ISSN 0920-5691
6. Miao Z, Jiang X (2013) Interest point detection using rank order log filter. Pattern Recognit 46(11):2890–2901
7. Criado R, Romance M, Sfanchez A (2012) Interest point detection in images using complex network analysis. J Comput Appl Math 236(12):2975–2980. ISSN 0377-0427
8. Ding G, Dai Q, Xu W, Yang F (2005) Affine-invariant image retrieval based on wavelet interest points. In: 2005 IEEE 7th workshop on multimedia signal processing, pp 1–4
9. Saydam SR, Rube IAE, Shoukry AA (2008) Contourlet based interest points detector. In: 2008 20th IEEE international conference on tools with artificial intelligence, vol 2, pp 509–513

10. Gevrekci M, Gunturk BK (2008) Reliable interest point detection under large illumination variations. In: 2008 15th IEEE international conference on image processing, pp 869–872
11. Maver J (2009) Self-similarity and points of interest. IEEE Trans Pattern Anal Mach Intell 32:1211–1226. ISSN 0162-8828
12. Lee WT, Chen HT (2009) Histogram-based interest point detectors. In: 2009 IEEE conference on computer vision and pattern recognition, pp 1590–1596
13. Martins P, de Carvalho P (2009) On interest point detection under a landmark-based medical image registration context. In: 2009 16th IEEE international conference on image processing (ICIP). IEEE, pp 2529–2532
14. Xie H, Gao K, Zhang Y, Li J, Liu Y (2010) GPU-based fast scale invariant interest point detector. In: 2010 IEEE international conference on acoustics, speech and signal processing, pp 2494–2497
15. Li B, Xiao R, Li Z, Cai R, Lu BL, Zhang L (2011) Rank-sift: learning to rank repeatable local interest points. CVPR 2011:1737–1744
16. Zukal M, Cika P (2012) Corner detectors: evaluation of information content. In: 2012 35th international conference on telecommunications and signal processing, pp 763–767
17. Wen H, Sheng XY (2011) An improved SIFT operator-based image registration using cross-correlation information. In: 2011 4th international congress on image and signal processing (CISP), vol 2. IEEE, pp 869–873
18. Yanhai W, Cheng Z, Jing W, Nan W (2015) Image registration method based on surf and freak. In: 2015 IEEE international conference on signal processing, communications and computing (ICSPCC), pp 1–4
19. Ma W, Wen Z, Wu Y, Jiao L, Gong M, Zheng Y, Liu L (2017) Remote sensing image registration with modified sift and enhanced feature matching. IEEE Geosci Remote Sens Lett 14(1):3–7. ISSN 1545-598X
20. Nasir H, Stankovic V, Marshall S (2010) Image registration for super resolution using scale invariant feature transform, belief propagation and random sampling consensus. In: 2010 18th European signal processing conference, pp 299–303
21. Qi X, Fuller E, Wu Q, Wu Y, Zhang CQ (2012) Laplacian centrality: a new centrality measure for weighted networks. Inf Sci 194:240–253
22. Qi X, Duval RD, Christensen K, Fuller E, Spahiu A, Wu Q, Wu Y, Tang W, Zhang C et al (2013) Terrorist networks, network energy and node removal: a new measure of centrality based on Laplacian energy. Soc Netw 2(01):19
23. P. N. Pournami and V. K. Govindan, "Interest point detection based on Laplacian energy of local image network," 2017 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET), Chennai, 2017, pp. 58–62
24. Schmid C, Mohr R, Bauckhage C (2000) Evaluation of interest point detectors. Int J Comput Vis 37(2):151–172
25. Oxford Dataset. http://www.robots.ox.ac.uk/~vgg/research/affine/

A Survey on Face Recognition in Video Surveillance

V. D. Ambeth Kumar, S. Ramya, H. Divakar and G. Kumutha Rajeswari

Abstract In today's world, enormous amount of threats arises due to terrorists, criminals, thieves and also illegal access of the data from the unwanted person, etc. This leads to a lot of challenges in our daily life. With the increase in threat globally the need to deploy reliable surveillance is to increase. Video surveillance is considered to be the major breakthrough in monitoring and security. In video surveillance, the facial recognition furthermore enhances the security and defense progressively. By face recognition the probe person can be recognized more accurately, efficiently and with short time. Various methods, approaches, algorithms were available for face recognition from surveillance video. The main objective of the paper is to discuss and analyze about the various facial recognition techniques.

1 Introduction

The concept of face recognition is to detect the particular character from the video or from the database and is used for security purposes. The face gives a rich source of information about a person and it is a most acceptable biometric. Face recognition, voice recognition, retinal scanning, fingerprint are some of the emerging biometric. In these, the method of spotting an individual by face recognition is far more accurate and faster than any others. Significant advances have occurred over the most recent couple of decades in face recognition. There is a significant attention and active research in this field. The human face tends to change significantly and quickly in

V. D. Ambeth Kumar (✉) · S. Ramya · H. Divakar · G. Kumutha Rajeswari
Computer Science and Engineering, Panimalar Engineering College, Chennai, India
e-mail: vdambethkumar@gmail.com

S. Ramya
e-mail: ramyasesha05@gmail.com

H. Divakar
e-mail: divakarvj06@gmail.com

G. Kumutha Rajeswari
e-mail: kums210819952@gmail.com

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_69

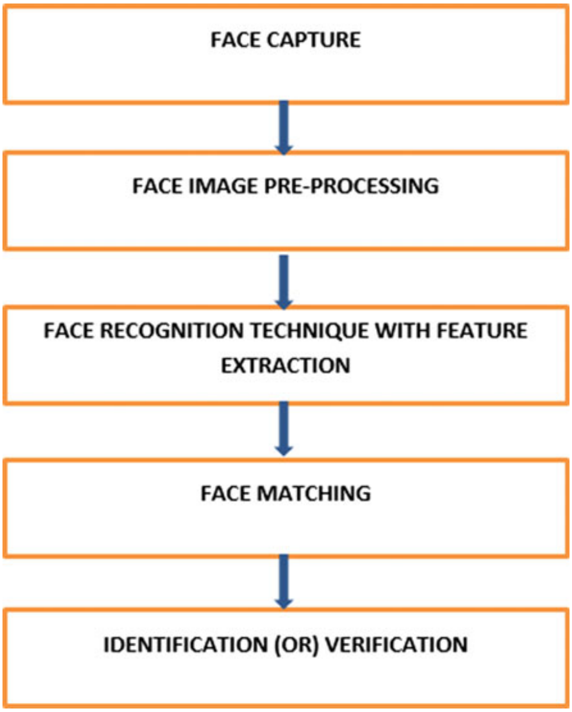


Fig. 1 Stage of face recognition algorithm

time, so it is considered as a complicated and dynamic structure. Face recognition is regarded to be challenging when there is variability in information due to pose variation, lighting conditions and different components. Face recognition should address these major challenges. Pose variation is considered to be a difficult problem because all faces seem to be similar with two eyes, mouth, nose and other features present in the same location.

The main aim of face recognition is to detect and verify a person from videos by one-to-many matching that compares the query face with the stored database faces [1]. The face recognition scheme may fail when there is a large difference in the query face and the stored database faces, e.g., makeup changes [2]. Face detection, feature extraction and face recognition are said to be the three stages of a face recognition system [3]. Face location is the way toward deciding the presence of a face in a video outline. Once a face is detected, face recognition is performed by isolating the face region and feature extraction is do carried out. Face verification and face identification are the two important stages of face recognition problem [4, 5].

The image preprocessing, feature extraction from face and the face matching are considered to be the main stages of the face recognition technique. In Fig. 1 these three main stages of face recognition algorithms are shown. Image normalization and enhancement are done during the image preprocessing. Also, the features of the face are extracted during this stage.

2 Face Detection Techniques

2.1 Conditional Random Fields

The Face Detection is done by dividing the face into various local regions optimally and integrating its multiple features and dependencies. These are demonstrated by a CRF structure. Multiple relationships modeling [6] are used for face detection in CRF.

2.2 VGG-Face Algorithm

Faces can be accurately recognized in spite of the difference in illumination using VGG-face algorithm [7]. Here, VGG-face performance is trialed with 8 NIST face recognition benchmarks. This algorithm performs exceptionally well in most difficult benchmarks.

2.3 Convolution Neural Networks

The CNN approach concentrates on the issue of face recognition in extreme pose variations. Different stances of particular models and rendered face images are utilized to handle posture variation [8]. This provides remarkably better performance. The CNN approach combined with a manifold based track comparison strategy can also be used for low-resolution video face recognition.

2.4 Gabor Feature-Based Local Generic Representation (G-LGR)

The G-LGR approach uses sparse representation properties for face recognition. This approach makes the recognition effectively, even when just a single probe subject for every class is feasible [9]. Feature extractions, such as uniform local binary patterns are not achieved by this technique.

2.5 Feature-Augmented and Feature Transformation

The technique of feature-augmented and feature transformation is combined in a hierarchical framework [10]. This multi-level component learning is utilized for confront acknowledgment under cosmetics changes. This is a robust method which is found to be superior to others. A level-wise change is utilized to limit the cosmetics and non-cosmetics confront contrasts.

2.6 Discriminant Correlation Analysis

DCA is one of the low-determination face detection strategies [11]. The low-resolution probe image is matched with high-resolution gallery face images. This computation is very much efficient and it can be applied to challenging real-time applications.

2.7 Heterogeneous Face Recognition

The heterogeneous face recognition focuses on the infrared-to-visible face matching scenario [12]. The main advantage of this technique is face recognition can be made possible in fog and haze and in low-light conditions. It uses an active illuminator which is not observable to the human eye. This method is purely passive and offers a truly covert surveillance capability.

2.8 Weber-Face and Singular Value Decomposition

This method Weber-Face and Singular Value Decomposition (SVD) is utilized to enhance the exactness of acknowledgment in a face acknowledgment framework. This technique can be utilized as a part of face acknowledgment in changing illumination [13]. At first, the face is represented by Weber–face (WF) method, and then the probe image is applied to singular value decomposition (SVD) method. Then, the SVD matrices value of test image is combined to adjust the illumination which is finally encoded by LBP/LTP descriptor.

2.9 Stacked Supervised Auto-Encoders (SSAE)

The SSAE method performs face identification task even though provided with only single sample per person (SSPP) in the gallery. This technique is additionally utilized for face detection under variety in posture, different facial expression and also in various illuminations [14]. Even though, the strategies combining with multiple samples per person probes (MSPPP) and SSAE will overcome the SSPP method.

2.10 Kernelized Locality-Sensitive Group Sparsity Representation

In paper [15], Face detection is performed by the joint sparse representation method. This sparse representation is the combination of group sparsity and locality-sensitive constraints. This improves the performance of recognition even in variation of pose. This KLS-GSRE method performs better than other sparse representation methods.

2.11 Discriminative Multimanifold Analysis for Face Recognition

Mostly the face recognition is done effectively by multiple samples per person (MSPP). But in this paper [16], a discriminative multimanifold analysis (DMMA) method is used for effectively to recognize face with single sample per person (SSPP).

2.12 Supervised Auto Encoder

The Supervised Auto encoder method should likewise be possible by utilizing single sample per person. This technique can extract facial features in various illumination changes, expression and occlusion, pose differences and facilitates face recognition [17]. Face verification can also be done by using this supervised autoencoder (Table 1).

3 Features of Human Face

The Description of the features of the face that are used in the face recognition system is discussed here. Some of the feature experts of the face are eye, chin, cheek, hair, jawline, mouth, nose, face outline etc. In these eyebrows, eyes, mouth, nose are considered to be the internal features while chin, face outline and hair are considered as external features. Although a system which uses an individual feature expert is good, the system which integrates these experts provides an even better result.

3.1 Eye

The Eye is the most widely used and attractive feature in face recognition. The features developed to detect eye is the variance projection function [18]. Eye detection

Table 1 Face recognition techniques

Name	Authors	Issues addressed	Methodologies used	Merits and demerits
A cross benchmark assessment of a deep convolutional neural network for face recognition	P. Jonathon Phillips	Face recognition in variable illumination	The VGG-face algorithm was trailed on eight NIST face recognition benchmarks	This algorithm performs exceptionally well in most difficult benchmarks It may not be possible to achieve optimal performance for all scenarios
Pose-aware face recognition in the wild	Iacopo Masi, Stephen Rawls, G´erard Medioni PremNataraja	Focus on the problem of extreme pose variations	Convolutional neural networks (CNNs)	Remarkably better performance
Gabor feature-based local generic representation for face recognition with single sample per person	Taher Khadhraoui	The primary problem tended to here is that if only one training subject per class is available	Gabor feature-based local generic representation (G-LGR) Virtual samples of each probe are considered and the new sample generic of a gallery set is used in order to generate the intrapersonal variations of different individuals	Have optimal localization properties in both spatial and frequency domains
Multi-level feature learning for face recognition under makeup changes	Zhenzhu Zheng and Chandra Kambhamettu	Propose a hierarchical framework to solve the problem of face recognition with cosmetic changes	Combine both strategies of feature augmentation and feature transformation in a unified way	A good representation is obtained by assembling multi-level representation. It is robust across makeup variations

(continued)

Table 1 (continued)

Name	Authors	Issues addressed	Methodologies used	Merits and demerits
Low-resolution face recognition in surveillance systems using discriminant correlation analysis	Mohammad Haghighat and Mohamed Abdel-Mottaleb	DCA analyzes the correlation of the features in high-resolution and low-resolution images and aims to find Projections that maximize the pair-wise correlations	A low-resolution face detection method based on Discriminant Correlation Analysis (DCA)	It has a very less computational Complexity. It can be used for real-time processing of several faces in a crowded image
Heterogeneous face recognition: recent advances in infrared-to-visible matching	Shuowen Hu1, Nathaniel Short, Benjamin S. Riggan, Matthew Chasse, M. Saquib Sarfraz	Matching between facial images acquired from different sensing modalities	Heterogeneous face recognition	Imaging through fog and haze, and in low-light conditions with an active illuminator not observable to the human eye heterogeneous face recognition is much more challenging
Five principles for crowd-source experiments in face recognition	Alice J. O'Toole, P. Jonathon Phillips	Crowd sourcing and human face identification on large databases. and to enhance the quality and accuracy of crowd-sourced data	Deep learning algorithms that consist of multi-layered neural networks	Accuracy and stability in face detection. It is critical to achieve human-sourced data that are meaningful and stable
Automatic face recognition.	Nawaf Yousef Al Mudhahka	The aim of semantic face recognition is to retrieve a suspect from a database of subjects using a human description of the suspect's face soft biometrics	Comparative soft biometrics is used due their ability for recognition and retrieval in constrained and unconstrained environments	Gives retrieval accuracy. Reduce dependency on human annotators. Bridges the semantic gap between humans and machines

(continued)

Table 1 (continued)

Name	Authors	Issues addressed	Methodologies used	Merits and demerits
Heterogeneous face recognition by margin-based cross-modality metric learning	Jing Huo, Yang Gao	Heterogeneous face recognition deals with matching face images from different modalities or sources	Margin-based cross-modality metric learning	Minimize intrapersonal cross-modality distances. Effective and superior
face recognition for movie character and actor discrimination based on similarity scores	Remigiusz Baran, Filip Rudzinski	A novel face detection approach dedicated to discriminate between motion picture characters and actors is presented in the paper	A bunching strategy in light of likeness scores ascertained by chose confront include descriptors is the key Element of this approach	Pretty high accuracy of face clustering up to 98%. Effectiveness strongly Depends on the accuracy of the eye localization process

is generally based on the eye motion and shape. The other method of face recognition using eye is Eye template matching [19]. The size of the image plays an important role in the calculation time. This method is considered to be efficient in terms of fast computation and also independent of the operating system platforms.

3.2 Eyebrows

Eyebrows are considered to be one of the prominent features in face recognition. The role of eyebrow in face recognition is considered to be as influential as that of the eyes. In the paper [20] eyebrows are used for detecting facial attractiveness and also to identify sexual dimorphism. An improved understanding of this system can contribute to the improvement of the artificial systems.

3.3 Skin Colors and Shape Information

The face recognition process is considered to be complex due to variation in illumination and background, difference in visual angle and facial expression. Once the skin color is evaluated, the shape information is used to locate the exact face. In the paper [21], the face candidates are located based on their color and shape information. Hue and saturation values are extracted and computed to find the best fit ellipse for each region. In the paper [22], facial regions are detected based on color and shape information. Here shapes are characterized by oval shape and HSV information is used for color localization. This is done by morphological operation and minima localization to intensity images.

3.4 Nose

A 3D nose tip localization method is used for face recognition. KNN AURA algorithm [23] is used to identify the nose tip, which is considered to be the facial feature to detect the faces. The identification rate of this method is 99.96% with much robustness and effectiveness.

The other method of face recognition using nose is by matching multiple overlapping regions around the nose [24]. This method is used for face recognition during varied facial expressions. This paper is considered to propose the first approach to solve expression variation problem.

4 Conclusion

In this paper, we have presented a review on face recognition. Different face recognition technique and the feature experts used for face recognition are discussed here. The three stages of the face recognition algorithm are image preprocessing, feature extraction and template matching. Obviously more accurate survey can be done by analyzing even more face recognition methods.

References

1. Patil SA, Deore PJ (2013) face recognition: a survey. Inform Eng Int J (IEIJ) 31–41
2. Zheng Z, Kambhamettu C (2017) Multi-level feature learning for face recognition under makeup changes. In: 12th International conference on automatic face and gesture recognition. IEEE, pp 918–923
3. Chihaoui M, Elkefi A, Bellil W, Ben Amar C (2016) A survey of 2D face recognition techniques. Computers 1–28
4. Meethongjan K, Mohamad D (2007) A summary of literature review: face recognition. In: Postgraduate annual research seminar, pp 1–12
5. Vijayakumari V (2013) face recognition techniques: a survey. World J Comput Appl Technol 41–50
6. Pang L, Ngo C-W (2015) Unsupervised celebrity face naming in web videos. IEEE Trans Multimed 17(6):854–856
7. Jonathon Phillips P (2017) A cross benchmark assessment of a deep convolutional neural network for face recognition. In: IEEE 12th international conference on automatic face and gesture recognition, pp 705–710

8. Masi I, Rawls S, Medioni G, Natarajan P (2015) Pose-aware face recognition in the wild. In: CVPR, pp 4838–4868
9. Khadhraoui T (2017) Gabor-feature based local generic representation for face recognition with single sample per person. IEEE, pp 157–160
10. Zheng Z, Kambhamettu C (2017) Multi-level feature learning for face recognition under makeup changes. In: IEEE 12th international conference on automatic face and gesture recognition, pp 918–933
11. Haghighat M, Abdel-Mottaleb M (2017) Low resolution face recognition in surveillance systems using discriminant correlation analysis. IEEE, pp 912–917
12. Hu S, Short N, Riggan BS, Chasse M, Sarfraz MS (2017) Heterogeneous face recognition: recent advances in infrared-to-visible matching. IEEE, pp 883–890
13. Tran C-K, Tseng C-D, Lee T-F (2016) Improving the face recognition accuracy under varying illumination conditions for local binary patterns and local ternary patterns based on weberface and singular value decomposition, pp 5–9
14. Vega PJS, Feitosa RQ, Quirita VHA, Happ PN (2016) Single sample face recognition from video via stacked supervised auto-encoder, pp 96–103
15. Tan S, Sun X, Chan W, Qu L, Shao L (2017) Robust face recognition with kernelized locality-sensitive group sparsity representation. IEEE, pp 1–8
16. Lu J, Tan Y-P, Wang G (2013) Discriminative multimanifold analysis for face recognition from a single training sample per person 1:39–51. IEEE
17. Gao S, Zhang Y, Jia K, Lu J, Zhang Y (2015) Single sample face recognition via learning deep supervised autoencoders. IEEE, pp 2108–2118
18. Feng GC, Yuen PC (1988) Variance projection function and its application to eye detection for human face recognition. Pattern Recognit Lett 19:899–906
19. Nikkhouy E, Abusham EEA (2011) Facial features detection using eyes-nose template. IJCSNS Int J Comput Sci Netw Secur 87–91
20. Sadrô J, Jarudi I, Sinha P (2003) The role of eyebrows in face recognition perception, pp 285–293
21. Sobotka K, Pitas I (1999) Extraction of facial regions and features using color and shape information. IEEE, pp 421–425
22. Sobottka K, Pitas I (1996) Face localization and facial feature extraction based on shape and color information. IEEE, pp 483–486
23. Ju Q (2013) Robust binary neural networks based 3D face detection and accurate faceregistration. Int J Comput Intell Syst 669–683
24. Chang KI, Bowyer KW, Flynn PJ (2006) Multiple nose region matching for 3D face recognition under varying facial expression. IEEE, pp 1695–1700

Driver's Drowsiness Detection Using Image Processing

Prajakta Gilbile, Pradnya Bhore, Amruta Kadam and Kshama Balbudhe

Abstract There are some causes of car accidents due to driver error which includes drunkenness, fatigue and drowsiness. Hence, the system is needed which will alert driver before he/she falls asleep and number of accidents can be reduced. In the proposed system, a camera continuously captures movement of the driver. To determine whether a driver is feeling drowsy or not the head position, eye closing duration and eye blink rate are used. Using this information, the drowsiness level is determined. As per the drowsiness level the alarm is generated. A night vision camera is used to handle different light conditions.

1 Introduction

Almost all the statistics have identified driver's drowsiness as a high priority vehicle safety issue [1]. Drowsiness refers to feeling sleepy, tired or being unable to keep eyes open. Due to drowsiness, the driver canot concentrate while driving, eye blink rate is decreased or increased and unable to keep eyes open. Fall-asleep crashes are very serious in terms of injury severity and may result in death. Drowsiness affects mental alertness and decrease an individual's capability to handle a vehicle safely. A driver is unable to predict when he or she will have an uncontrolled sleep onset. The advancement in technologies develops interest in driver's safety and comfort, increase traffic flow and reduce accidents. This paper introduces an alerting process when the driver falls asleep. It calculates the level of drowsiness depending on the

P. Gilbile (✉) · P. Bhore · A. Kadam · K. Balbudhe
Department of Information Technology, PVG's COET, Pune, India
e-mail: prajakta.u.gilbile136@gmail.com

P. Bhore
e-mail: pradnyab208@gmail.com

A. Kadam
e-mail: amrutakadam999@gmail.com

K. Balbudhe
e-mail: ksb_it@pvgcoet.ac.in

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_70

head position, eye blinking rate. If the level exceeds the limit from a threshold, then the alarm is generated. Different sound alarms such as ‘Take a break’ or ‘Have a coffee’, etc., must be given at particular level of drowsiness. So that before falling asleep driver will get alert.

2 Related Work

Khunisuth et al. [2] discussed paper in which the drowsiness is detected based on various factors such as titling of head, blinking of eyes and eye blink rate [3]. Image is captured through camera then localization of head is done [4]. It is followed by localization of eyes and then titling of head angle is detected. Different templates such as both eyes closed, right eye closed, left eye closed, both eyes open are used to detect eyes. A drawback of this paper is that it is not steady in all light conditions. The accuracy of 99.59% is achieved for only stable light conditions. In the proposed system, the night vision camera is used to handle different light conditions.

Ahmed et al. [5] discussed paper in which drowsiness is detected based on only eyes. First, the location of driver’s eyes is located and then it is decided whether the eyes are open or not. The captured image is in binary form through which location of edges of face are detected. This ultimately gives the location of eyes. Various frames are captured. If the driver’s eyes found to be close in five successive frames, then the system assumes that drives is feeling sleepy and alert gets generated accordingly. The drawback of this paper is that, it may happen that some people have habit of blinking of eyes more than that the normal rate. The result which will get generated will be ultimately wrong. Only eyes cannot give accurate output to determine drowsiness.

Tadesse et al. [6] discussed paper which mainly focuses on the facial expressions for drowsiness detection. Many previous papers are specifically focus on eye closure and blinking of the driver. The facial expressions of driver are analysed through Hidden Markov Model (HMM-based dynamic modelling). Every time presetting of window size to fixed value is needed for different parameters.

Abtahi et al. [7] discussed a paper in which drowsiness is detected on basis of yawing of driver. A camera detects the face and eyes of the driver. After detection of eye, mouth is detected and then successively yawning is detected. On basis of only yawing, alert is generated. Yawing is detected in two steps. In first step yawn component is detected and in second step mouth location is used to verify the validity of detected component. This paper uses number of algorithms so they are insensitive to changes in light conditions, skin types. Various verification techniques are used to reduce false rate. The limitation of this paper is that it detects the drowsiness on basis of only yawning and uses number algorithms.

Saini et al. [8] discussed a way to find drowsiness which uses ECG and EEG, Steering Wheel Movement (SWM), Local Binary Pattern (LBP), and Optical Detection. This paper also discusses about eye blink-based technique, head nodding, and yawning-based technique.

3 Proposed Work

The main concept of 'Driver's Drowsiness Detection' is to capture a driver's face using a camera and accurately calculate the level of drowsiness. The proposed system consists of a camera pointing at the driver. A camera continuously captures images of driver. There are main five stages of processing: The first stage is to capture image using camera. Second stage is localization of head and check head position. Third stage is calculation of eye blink rate. Fourth stage is calculation of eye closing duration and fifth stage is to generate the alert. At different levels of drowsiness, different alerts will get generated.

3.1 Capture Images from Camera

The camera is used for continuously capturing video. The camera is placed such a way so that it will capture all head movements and eye movements of driver. Due to different light conditions the noise can be introduce in image. To handle different light conditions including night time, the night vision camera can be used. Camera will give whole picture of driver with background details but we are interested only in head position area. JMyron and OpenCV library is used to capture image and finding out area of interest. The captured image is RGB image and it is transformed into grey scale image for processing. As image is stored temporarily and after processing it is discarded, it does not take large space.

3.2 Find Head Position

For finding head position (Area of Interest) 'Haar-Cascade' algorithm is used [9]. JMyron library will remove unnecessary portion (background details) from captured image and give the area in which head is present. On which 'Haar-Cascade' algorithm is applied. As mentioned in algorithm JMyron library returns a vertex point of frame (ROI) in which head is present. Using this point, height and width of frame, the centre point of frame is calculated which is ultimately the centre point of head. The standard centre point is already defined to indicate position of head when driver is not feeling drowsy. The centre point of new frame is compared with the standard centre point and the difference between these two points is calculated. If the difference is greater than threshold value then alarm is generated. By using the difference, a level of drowsiness can be determined. If the difference is greater then, the drowsiness level is high and if low then, the drowsiness level is low. The threshold value is approximately 100 or 150 pixels.

3.3 Eye Blink Rate Calculation

After localization of eye, eye blink rate is calculated. The average eye blink rate of human beings is approximately fifteen to twenty times in one minute. For comparing with current eye blink rate two values for threshold are set.

High_threshold_binkrate = 25
Low_threshold_binkrate = 10

System captures eye blink rate for one minute. If current eye blink rate is greater than high_threshold_value or less than low_threshold_value then alarm is generated. Template matching is not used to detect eyes open or not. If eye blink rate is too fast or low then, drowsiness level is high.

3.4 Eye Closing Time Calculation

System captures eye closing time by checking successive frame. If eye closing time is greater than the threshold value, alarm is generated. The approximate threshold value is 3s.

3.5 Alarm Generation

This module is responsible to alert driver. Depending upon level of drowsiness different alarms are generated. This can be buzzer or voice message such as 'You need a break', 'Take a cup of coffee' etc. Another way is to include hardware like vibrator to alert driver.

4 Algorithm

Let P (xF, yF) = (0,0) be the vertex point of the captured frame. Let P (standX, standY) be the standard centre point for head position. Let fRect be the Region of Interest in which head is present.

Algorithm Drowsiness detection ()

{

Apply Haar Cascade algorithm to get vertex point P (xSelect, ySelect) of fRect.

$$currCX = ((xF + xSelect + fRect.x()) + fRect.width()/2) \tag{1}$$

$$currCY = ((yF + ySelect + fRect.y()) + fRect.height()/2) \tag{2}$$

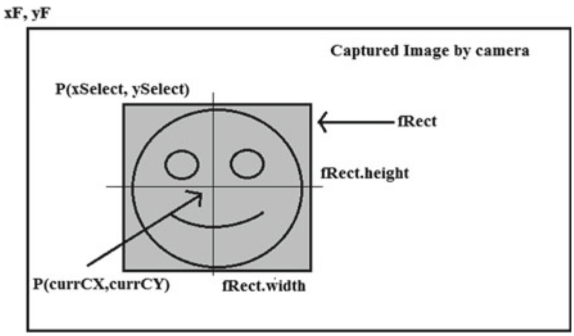


Fig. 1 Terminologies used in algorithm

Find the distance between P (standX, standY) and P (currCX, currCY)

$$Distance = \sqrt{(currCX - standX) * (currCX - standX) + (currCY - standY) * (currCY - standY)} \tag{3}$$

If (Distance > threshold value of head position)
Generate alarm.

$$wF = fRect.width() \tag{4}$$

$$hF = fRect.height() \tag{5}$$

Select (0, 0, wF/2, hF) and (wF/2, 0, wF/2, hF) as Rect1 for left eye and Rect2 for right eye, respectively.

Apply Haar Cascade algorithm on Rect1 and Rect2 to detect eyes.

If (Rect1 == null && Rect2 == null)

Blink_count ++;

If (Blink_count > high threshold value of eye blink rate || Blink_count < low threshold value of eye blink rate)

Generate the alarm.

If (Rect1 == null && Rect2 == null) for continuous 3 s
Generate the alarm.

} (Fig. 1)

5 Conclusion

In this paper, the way to identify level of drowsiness is specified and according to that different alarms can be given to driver. The night vision camera is used to handle different light conditions. The algorithm to find head position is specified. In algorithm, the drowsiness level is determined depending upon head position, eye blink rate and eye closing duration. Proposed system detects drowsiness level and helps a driver to stay awake while driving. This system will reduce car accidents about great extent.

References

1. Ahmad R, Borole JN (2015) Drowsy driver identification using eye blink detection. Int J Comput Sci Inf Technol. 6(1):270–274
2. Khunpisuth O, Chotchinasri T, Koschakosai V, Hnoohom N (2016) Driver drowsiness detection using eye-closeness detection In: Signal-Image Technology & Internet-Based Systems (SITIS), 2016 12th International Conference on, pp. 661–668. IEEE, 2016
3. Parmar SH, Jajal M, Brijbhan YP (2014) Drowsy driver warning system using image processing. Int J Eng Dev Res, IJEDR1303017
4. Kuo Y-C, Hsu W-L (2010) Real-time drowsiness detection system for intelligent vehicles. Proceedings of the 5th Symposium on Smart Life Science and Technology (Part 1)
5. Ahmed J, Li J-P, Khan SA, Shaikh RA (2015) Eye behavior based drowsiness detection system In: Wavelet Active Media Technology and Information Processing (ICCWAMTIP) 2015 12th International Computer Conference on, pp. 268–272. IEEE, 2015
6. Tadesse E, Sheng W, Liu M (2014) Driver drowsiness detection through hmm based dynamic modelling In: Robotics and Automation (ICRA) 2014 IEEE international conference on robotics and automation (ICRA), pp. 4003–4008. IEEE, 2014
7. Abtahi S, Hariri B, Shirmohammadi S (2011) Driver drowsiness monitoring based on yawning detection In: Instrumentation and Measurement Technology Conference (I2MTC), pp. 1–4. IEEE, 2011
8. Saini V, Saini R (2014) Driver drowsiness detection system and techniques: a review. Int J Comput Sci Inf Technol. 5(3):4245–4249
9. Pamnani R, Siddiqui F, Gajara D, Gupta A, Pandya K Driver drowsiness detection using haar classifier and template matching. Int J Adv Res Eng Technol 3(IV), April ISSN 2320–6802
10. Nguyen TP, Chew MT, Demidenko S (2015) Eye tracking system to detect driver drowsiness In: Automation, Robotics and Applications (ICARA), 2015 6th International Conference on, pp. 472–477. IEEE, 2015
11. Assari MA, Rahmati M (2011) Driver drowsiness detection using face expression recognition In: Signal and Image Processing Applications (ICSIPA), 2011 IEEE International Conference on, pp. 337–341. IEEE, 2011
12. Flores MJ, Armingol JM, de la Escalera A (2010) Real-time warning system for driver drowsiness detection using visual information. J Intell Robot Syst 59(2):103–125

Identification of Optimum Image Capturing Technique for Corneal Segmentation—A Survey

H. James Deva Koresh and Shanty Chacko

Abstract Segmentation of corneal layers plays an important role in diagnosing the corneal disease and it also helpful for planning the refractive surgeries when measuring the corneal layer thicknesses. There are several imaging instruments available for capturing the image of the cornea which can be used for segmentation and thickness measurement of corneal layers. The work describes the various available imaging instruments along with the necessity of imaging the cornea. It also suggests a reliable imaging method for capturing the cornea for segmentation by doing a comparative study. It will be helpful for identifying the best image capture method for implementing the image processing technique for further segmentation and measurement flow from the image.

1 Introduction

The refractive surgeries like laser-assisted in situ keratomileusis (LASIK) and photorefractive keratectomy (PRK) which are the methods used to correct minor and modular refractive errors till −8.00 for myopia and <+4.00 for hyperopia. The corneal segments consist of Epithelium, Bowman's membrane, Stroma, Descemet's membrane, and the Endothelium. The typical corneal thickness of a human eye is 550 µm at the center and it gradually increases towards the marginal area [1, 2]. The ideal thickness of the corneal layer may vary due to changes in oxygen consumption of cornea, acidosis and swelling. So, the accuracy in evaluation of corneal thickness is necessary to diagnose such problems to provide proper treatment [3, 4].

H. James Deva Koresh (✉)
Department of Electronics and Communication Engineering,
Karunya Institute of Technology and Sciences, Coimbatore, India
e-mail: jamesdeva@karunya.edu.in

S. Chacko
Department of Electrical and Electronics Engineering,
Karunya Institute of Technology and Sciences, Coimbatore, India
e-mail: shanty@karunya.edu

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_71

The epithelium is the outermost layer of the cornea which covers less than 10% of the central corneal thickness (CCT), the thickness is almost around 50 μm for a healthy eye. The tissues in the epithelium layer have the capability of replacing itself after a damaged (in LASIK) or removal (in PRK). The regeneration time of epithelium layer is less than a week depends upon the changes. The Bowman's layer is a non-regenerative connecting tissue of 8–14 μm lies between epithelium and stroma. The Bowman's layer is ablated in the photorefractive keratectomy (LASIK or PRK). The stroma is a deeper layer that covers 90% thickness of cornea with a thickness of 500 μm. The stroma is contrary to epithelium consists of permanent tissues which will remain permanently after a reshape. The Descemet's membrane of 15 μm depth is the next layer to stroma which separates the stroma layer from endothelium. This membrane gets increases from 5 μm in children to 15 μm of an adult. The endothelium is a single-layer cell of 5 μm thickness and it is the innermost layer of the cornea [2].

In LASIK the epithelium layer is damaged in a circular manner to create a flap with the help of microkeratome or femtosecond laser to access further layers and the flap is laid back to position after the procedure to heal [5]. For PRK the epithelium layer is removed to access further layers. In both LASIK and PRK a cool ultraviolet light beam called laser is passed over the stroma to reshape it permanently. For myopia the laser beam is used to flatten the stroma layer and it makes steeper for hyperopia [5]. The corneal thickness must not be less than 470 μm for LASIK and PRK procedures and the final corneal thickness should be more than 400 μm after the procedures to have a healthy cornea [6]. The ablation depth of the stroma layer varies depends upon the refractive error of an eye and the percentage of tissues to be altered (PTA) should not exceed 40% of the CCT [6]. In refractive surgery the thinnest cornea may results in ectasia, it is an abnormal bulging over the surface of the cornea [7]. Hence it is must to measure the CCT of an eye before the photorefractive keratectomy (LASIK and PRK) to analyze the PTA.

2 Related Work

There are several techniques are available for estimating the corneal thickness. Pachymeter is a common name for devices which are used to measure the corneal thickness and the procedure for estimating the corneal thickness is called as pachymetry. In general, the pachymeter devices are broadly classified as contact and non-contact types.

2.1 Ultrasound Pachymetry

The Ultrasound Pachymetry (UP) is a simple, cost-effective and handy device used in medical field, for estimating the CCT. The UP is a contact-type pachymeter has an ultrasound probe which is to be kept directly over the corneal apex to make an ultrasound waves. The probe helps us to find out the CCT with the help of receiving the reflected sound from the surfaces of cornea. The reflected sounds are further taken into the connected device through the probe as multiple readings for the estimation of average CCT and standard deviation among the readings. The user also has the accessible to view each measured values individually and can delete the value which seems to be too irrelevant to other readings. As the UP is a contact-type device, it requires anesthetic drops for the cornea to avoid itchiness. Due to the contact of the probe in direct to the cornea, there is also chance for damage and abrasion over the cornea. Hence, this method requires trained practitioner for utilizing this device and there are also chances for variation in readings due to misalignment of the probe by the practitioner. This misalignment may happen due to keeping the probe just closer to the central cornea, it will result in showing higher corneal thickness than actual. Also, the ultrasound probe should be kept 90° perpendicular to the central cornea. Small changes in keeping the probe may also lead to a false estimation of CCT [8]. From this device, the intraocular pressure (IOP) is also measured as it has a direct relationship with the CCT. The ultrasound pachymetry is also called as the perimeter.

2.2 Orbscan

The Orbscan corneal topography is a quick and non-contact method of estimating corneal thickness along with anterior and posterior surface evaluation. This scan provides the overall health of a cornea including corneal shape and the result of this procedure will be 3D maps of the surface of an eye [9, 10]. The Orbscan is working on the principle of slit-scanning technology and video keratography estimates the thickness by measuring the reflections of rays from the corneal surface [11]. The Orbscan delivers a pair of vertical scans at 45° angle from both sides of the cornea to make it as slits. The slits are analyzed further with the help of digital video camera to make a topography of the eye in 3D maps [12]. As it is a non-contact method it does not require anesthetic drops during the procedure.

2.3 Optical Coherence Tomography

Optical Coherence Tomography (OCT) is another non-invasive method that can be used for imaging the corneal and retinal segments with the help of infrared light. The OCT is used to examine 2–3 mm of subsurface tissues in retina and cornea [13, 14]. The OCT performs the scanning of the corneal area with infrared spots produced by Michelson-type interferometer. The OCT was used to analyze the retinal layer by passing the infrared spot over the eye with the wavelength of 1310 nm. The cross-sectional images of the corneal layer can be viewed in this process by passing the infrared spots with 820 nm wavelength over the corneal surface in multiple axes [11].

The image also has the cross-sectional view of anterior and posterior segments of an eye. The anterior segment imaging is helpful for analyzing the space available for implanting the intraocular lens (IOL) to correct ew32 the high refractive error of an eye from the range of >−8.00 for myopia (nearsightedness) and >+4.00 for hyperopia (farsightedness) [15]. The cross-section image of cornea provides the curvature and thickness of central to peripheral corneal layer, which is also useful in diagnostic corneal edema and ocular hypertension.

2.4 Pentacam

The Pentacam is a device used for imaging the corneal surface and anterior segments of an eye. The Pentacam is a fast imaging instrument requires only 2 s for capturing; this is achieved with the help of scheimpflug camera by rotating it for 180° on the optical axis along with a monochromatic slit light source. It will take around 50 slit images consists of 500 assessing points for further estimation of CCT and each corneal layers [1, 16]. The Pentacam is also known as Scheimpflug Tomography.

3 Literature Survey

Martínez-Albert et al. [17] conducted the repeatability test on OCT in measuring the CCT. The test was repeated three times on 30 healthy individuals on their right eye. The result indicates that there is better correlation between the reading measured on each time by the OCT and it provides the intraclass correlation coefficient (ICC) of 0.984–0.988 [17]. Otchere et al. [16] assessed the CCT from 20 keratoconus individuals with OCT and Pentacam for analyzing their repeatability and reproducibility. The CCT was measured twice in a day from each instrument and repeated the test after 48 h in same manner. The assessed CCT values explores that both the devices give good repeatability on all the surfaces over the cornea but the reproducibility among the two devices are not good enough [16]. Ramesh et al. [18] compared the CCT measurements taken from 120 normal corneas by UP and OCT. The readings from OCT were measured twice and 25 times from UP. The values are average for taking comparison and the result shows that the values observed by OCT are slightly high always comparing to the values of UP. Upon the high repeatability the OCT stands better than UP [18].

Antonios et al. [19] assessed the CCT for 44 eyes of Keratoconus in before and after the corneal cross-linking (CXL) procedure in the time span of 14, 30, 90, 180, and 360 days. The CCT has been assessed with UP, OCT, and Scheimpflug Tomography (Pentacam). It is observed from the mean CCT assessment that the OCT has made similar measurement obtained by the UP throughout the assessment period [19]. Xu et al. [20] made the reliability test on Pentacam by means of measuring the corneal thickness for 60 Healthy eyes, 60 post-LASIK eyes, and 27 Keratoconus eyes.

The repeatability and reproducibility of the Pentacam are analyzed by allowing two trained operators to measure corneal thickness for three times of an eye. The CCT shows best repeatability and reproducibility than the peripheral corneal thickness and it also stated that measuring the peripheral cornea thickness through Pentacam should be taken care of unhealthy cornea and eyes [20].

Temstet et al. [21] made a test for finding the thickness of the outermost layer of the cornea called epithelium with the help OCT to measure its detection ability over Orbscan and Pentacam. There are 145 eyes consists of healthy, fruste keratoconus, moderate keratoconus, and severe keratoconus were undergone for the test and it explores that the subjects with severe keratoconus are having thinnest epithelium layer and it was found perfectly with the help of OCT than other methods [21]. Kuerten et al. [22] performed comparison study on OCT, Pentacam, and UP by analyzing the repeatability and reproducibility of estimating the CCT among control and post corneal edema. The estimation was done by two individual investigators randomly. The test started Pentacam, OCT and finally, with UP, the UP has the highest nature of user dependence for analyzing the CCT [22]. Randleman et al. [23] estimated the CCT among 100 eyes for comparison of performance between OCT, Orbscan, Pentacam, and UP. The test result indicates that there are significant variations among the results observed by all the instruments and stated that the observed values from any of the instrument should not be applicable for another instrument in comparison.

Lin et al. [24] evaluated the repeatability of OCT with UP by estimating the CCT for 51 patients with glaucoma, suspected glaucoma, and cataract with two different examiners. First, the CCT measurement was taken with OCT and then with UP, each device is repeated for two times of estimation. The repeatability test was done by analyzing the observed CCT value in the intraclass correlation of coefficient (ICC) between UP and OCT. The result indicates high repeatability range of ≥ 0.978 and for 2 mm from the center, the ICC increased to ≥ 0.999 [24]. Yazici et al. [25] made the CCT test for 100 eyes with the single operator through OCT, Orbscan and Pentacam. The comparative study shows that OCT measured the values of CCT thinner to other two methods but the variations among these methods are small and it does not affect the decision for photorefractive keratectomy [25]. Doors et al. [26] did a comparative study of CCT measurement with 66 healthy eyes and 42 eyes of after intraocular lens implantation (IOL). The comparative study takes place between the non-contact methods of OCT, Orbscan, and Pentacam. The conclusion indicates that there is a substantial variation among the values calculated by all three methods and the study recommends to not use these devices interchangeably for analysis [26].

Cheng et al. [27] measured the CCT with UP, Orbscan, and OCT for 68 eyes of 34 myopia individuals after 6 months of their LASIK surgery. The measured CCT were compared by the linear regression method, paired sample t-test, Bland and Altman plot and the result shows a closed correlation between OCT and UP comparing with Orbscan [27]. Kim et al. [28] done a relative test on CCT for 155 individuals with the clinically healthy eye. Two measurements from OCT and UP were taken and averaged for one eye randomly from a patient. The reproductive difference in CCT is significantly present on both UP and OCT and the test suggests that these two methods should not be interchanged in clinical practice for analysis and measurements [28].

Li et al. [29] estimated the value of CCT for 70 eyes with OCT, UP, and Orbscan. The OCT underestimates the CCT among the other two methods and the OCT has a better agreement than Orbscan with UP. As the CCT measurement is highly correlated, it is just to avoid taking measurement interchangeably in clinical practices [29]. Ho et al. [30] evaluated the CCT for 103 post-LASIK individuals with OCT, Orbscan, Pentacam, and UP. The evaluation has taken place after the period of six months from the LASIK procedure and the comparative evaluation specifies that OCT and Pentacam attained thinner values than the rest of the methods but it has a great correlation between all four methods [30]. Zhao et al. [31] tested the CCT with OCT and UP to 285 normal individuals. The observed CCT values between OCT and UP had a very close relationship of Pearson correlation coefficient of 0.98, $P < 0.001$. The Bland Altman analysis shows that the CCT of UP is comparatively high by 16.5 ± 11.7 μm [31].

Haque et al. [32] assessed the CCT along with epithelium and stromal thickness of 20 keratoconus and 20 with normal individuals with the help of OCT, Orbscan, and UP. The assessed values of OCT and Orbscan have closer association than UP. The UP estimated the CCT as higher in both central apexes of the cornea. It also indicates that the keratoconus subjects have thinner cornea layers than the normal individuals [32]. Leung et al. [33] performed the CCT test for 50 healthy individuals with OCT and then with UP, the correlation and Bland Altman comparison indicate a high correlation between the OCT and UP with Pearson coefficient of 0.934 and in Bland Altman plot the OCT results are in higher range continuously by mean of 23 μm. The OCT is agreed with UP with minor adjustment factor [33]. Li et al. [34] analyzed the CCT of LASIK undergoing patients of 21 individuals. The analysis was taken by UP and OCT procedures and the UP shows the higher value of CCT in Pre-Lasik phase and lower value than OCT on Post-Lasik phase. The OCT and UP techniques are highly correlated by Pearson correlation of 0.98 in Post-LASIK and 0.97 for Pre-LASIK [34].

Lackner et al. [35] conducted and compared the CCT measurement with 30 healthy eyes for twice the time with two independent observers. The comparative study was done by taking standard deviation among the values and it shows that Pentacam shows the values closer to UP than Orbscan and the reproducibility of Pentacam is also comparatively high [35]. Fishman et al. [36] made an assessment of CCT with OCT, Orbscan, and UP for 22 eyes. The result of the assessment states that the OCT is a good reproductive and accurate non-invasive method for estimating the CCT. The correlation of CCT between of OCT is 0.984 and UP is 0.981 where the Orbscan has the correlation of 0.942 [36]. Wong et al. [37] had performed the CCT test with OCT, Orbscan, and UP for 74 eyes. The test was first carried out with the non-contact methods (OCT and Orbscan) following to the UP. There are five readings were taken for non-contact methods and three for UP. The average values were comparatively analyzed using correlation, linear regression, and one-way analysis of variance methods and the result shows that OCT is more accurate than Orbscan and UP [37].

4 Discussion

Table 1 explores the CCT values measured by various experts with different imaging instruments in healthy and unhealthy eyes. It also indicates that the OCT method is an optimum technique for capturing the image of cornea. Hence the OCT is further compared individually with Pentacam, Orbscan, and UP to get a better solution.

4.1 Pentacam Versus OCT

Lackner et al. states that in comparing the values of non-invasive method with invasive method, the Pentacam values are very closer to the UP and the reproducibility of Pentacam stands better than the other non-invasive method. Otchere et al. made a test to find the repeatability of Pentacam and OCT by measuring the reading twice in two days span. The result indicates that the Pentacam is not the only device with better repeatability as stated by Lacknet et al. It also indicates that OCT has the good repeatability with lesser range difference than Pentacam. Ho et al. result explores that the correlation between the non-invasive and invasive methods are compromising to each other with higher values but the observed values of CCT from Pentacam and OCT are comparatively lesser than the other methods. Antonios et al. prove that the CCT values obtained by OCT and UP have better similarity than the Pentacam by readings observed from the individuals after multiple times. Xu et al. warns to take more care in measuring the peripheral thickness of cornea through Pentacam for the operated cornea and illness cornea.

4.2 Orbscan Versus OCT

Li et al. observed that the difference between the values of CCT are lesser in Orbscan with UP than OCT. Also the difference is very small in Orbscan with UP. The difference among the values observed by OCT is comparatively higher when comparing with UP and Orbscan and the difference with both the devices are almost equal with certain variations in microns. Cheng et al. suggest that both Orbscan and OCT has the ability to measure thinner corneas accurately than UP. The obtained values from Orbscan and OCT are having better correlation together. Fishman et al. state that the OCT is an accurate method in measuring the values of CCT. The study shows that the range values obtained from several individuals are higher in Orbscan than OCT in comparing with UP.

Table 1 Comparison of CCT values measured by various experts with different imaging instruments in control and uncontrolled eyes

Author name	Measured parameters
Martínez-Albert et al. [17]	Mean CCT + SD in microns 520 ± 30.59
Otchere and Sorbara [16]	Mean and SD of CCT in microns OCT = 484.97 |±| 43.14; range: 484.84–486.09 Pentacam = 478.86 |±| 45.31; range: 477.20–480.53
Ramesh et al. [18]	Mean and SD of CCT in the right eye in microns OCT = 516.28 ± 29.76; UP = 532.42 ± 29.71 Mean and SD of CCT in the left eye in microns OCT = 515.82 ± 29.88; UP = 532.36 ± 29.83
Antonios et al. [19]	Mean CCT in microns OCT = 470.02; UP = 469.79; Pentacam = 466.66
Xu et al. [20]	Coefficient of variation of Central and Peripheral corneal thickness in all 3 states, Healthy = ≤ 1.3%; Post-LASIK = ≤ 1.6%; Keratoconus = ≤ 1.6%
Temstet et al. [21]	Epithelium thickness in microns for Form fruste keratoconus = 52.8 + 3.3; Moderate keratoconus = 49.5 + 5.2; Severe keratoconus = 47.9 + 4.9; Control = 53.0 + 3.1
Kuerten et al. [22]	CCT from investigator 1 Control Pentacam = 561.5 ± 38.64; OCT = 552.8 ± 38.73; UP = 547.6 ± 39.81 Post corneal edema Pentacam = 615.9 ± 58.02; OCT = 608.8 ± 65.67; UP = 601.4 ± 63.77 CCT from Investigator 2 Control Pentacam = 560.7 ± 38.56; OCT = 553.1 ± 38.39; UP = 554.4 ± 4 Post corneal edema Pentacam = 615.1 ± 60.17; OCT = 606.9 ± 64.41; UP = 614.5 ± 70.91
Randleman et al. [23]	Mean and SD of CCT in microns OCT = 550.5 ± 32.7; Orbscan = 570.9 ± 36.1; Pentacam = 552.8 ± 33.8; UP = 563.9 ± 36.1
Lin et al. [24]	Corneal thickness with standard deviation measured by examiner 1 at Central = 554.43 (30.42); vertex = 540.22 (30.00); central 2 mm = 540.86 (29.81) Corneal thickness with standard deviation measured by examiner 2 at Central = 551.39 (29.85); vertex = 541.24 (29.70); central 2 mm = 540.04 (29.54)
Yazici et al. [25]	OCT = 529 ± 30.5 μm; Orbscan = 554 ± 32.7 μm; Pentacam = 552 ± 29.3 μm
Doors et al. [26]	For healthy eyes, OCT and Orbscan ($P = 0.422$) For IOL eyes, Orbscan and Pentacam ($P = 0.214$)

(continued)

Table 1 (continued)

Author name	Measured parameters
Cheng et al. [27]	Average CCT of post-LASIK under UP = 436.65 ± 43.82 µm; Orbscan = 422.84 ± 51.04 µm; OCT = 422.26 ± 42.46 µm
Kim et al. [28]	Mean CCT + SD in microns UP = 523.3 ± 33.5; OCT = 499.0 ± 32.0 CCT range in microns UP = 422–653; OCT = 428–613
Li et al. [29]	UP = 553.5 ± 30.26 µm; Orbscan = 553.22 ± 25.47 µm; OCT = 538.79 ± 26.22 µm
Ho et al. [30]	The mean CCT values UP = 438.2 ± 41.18 µm; Orbscan = 435.17 ± 49.63 µm; Pentacam = 430.66 ± 40.23 µm and OCT = 426.56 ± 41.6 µm
Zhao et al. [31]	Pearson coefficient = 0.93, $P < 0.001$ Bland Altman = 16.5 ± 11.7 µm
Haque et al. [32]	CCT of Keratoconus UP = 494.2 ± 50.0 µm; Orbscan = 438.6 ± 47.7 µm; OCT = 433.5 ± 39.7 µm
Leung et al. [33]	Pearson coefficient = 0.934 Bland Altmon = 23 µm
Li et al. [34]	Before LASIK mean CCT + SD in microns UP = 546.9 ± 29.4; OCT = 553.3 ± 33.0 After LASIK mean CCT + SD in microns UP = 498 ± 46.6; OCT = 513.7 ± 44.5
Lackner et al. [35]	Pentacam = 542 ± 29 µm; Orbscan = 576 ± 37 µm; OCT = 530 ± 34 µm; UP = 552 ± 32 µm
Fishman et al. [36]	Mean CCT + SD in microns UP = 545.1 ± 37.6; OCT = 545.1 ± 36.8; Orbscan = 573.9 ± 51.4 CCT range in microns UP = 479–609; OCT = 481–614; Orbscan = 512–656
Wong et al. [37]	Orbscan = 555.96 ± 32.41 µm; UP = 555.11 ± 35.30 µm; OCT = 523.2 ± 33.54 µm

4.3 Ultrasound Pachymetry Versus OCT

Zhao et al. found that the CCT values observed from OCT constantly lesser than UP with substantial correlation. Kim et al. also found that there is always a significant difference between the values obtained by OCT in comparing with UP but Leung et al. prove that the OCT is an agreeable device with UP with minor adjustment factor. Li et al. result indicates the high correlation between OCT and UP on pre- and post-LASIK individuals. Also Fishman et al. suggest OCT method than Orbscan for its better accuracy. The OCT also has the ability to measures thinner corneas Temstet et al. Upon the repeatability of OCT device Ramesh et al., Martínez-Albert et al., Otchere et al. state that there is no massive change in the obtained readings and in comparing with Orbscan the OCT has the better repeatability. Otchere et al.

also mention the OCT as a device with good repeatability as like Pentacam but the range of values obtained by OCT is similar to the UP than Pentacam.

5 Conclusion

The values of CCT obtained by OCT, Orbscan, and UP can be equal with correlation factor of 32 μm [17]. The differences between the values observed by the non-invasive devices are not going to affect the refractive surgeries [18] as the differences are very few microns range. Also if there are differences in the readings among all the devices the correlation among them is in higher range [21]. For analyzing the CCT the non-invasive methods are not to be interchanged and compared to each other for assessment on before and after refractive surgery [1, 23]. Among the all three non-invasive methods the OCT have the high correlation with UP also the accuracy and reliability is also high in OCT when measuring the thickness in central and apex of the cornea [33]. The OCT is correlated with UP but the UP is an invasive method so the readings are user-dependent [34]. So the OCT is the optimum image capturing technique for segmentation of the corneal layer and thickness measurement in central and peripheral area of cornea with high repeatability.

References

1. Rabsilber TM, Khoramnia R, Auffarth GU (2006) Anterior chamber measurements using Pentacam rotating Scheimpflug camera. J Cataract Refract Surg 32(3):456–459
2. Gary H (2017) Cornea of the eye—definition and detailed illustration. All about vision. www.allaboutvision.com/resources/cornea.htm
3. DelMonte DW, Kim T (2011) Anatomy and physiology of the cornea. J Cataract Refract Surg 37(3):588–598
4. Eichel J, Mishra A, Fieguth P, Clausi D, Bizheva K (2009) A novel algorithm for extraction of the layers of the cornea. In: Canadian conference on computer and robot vision, CRV'09, pp 313–320. IEEE
5. Alila Medical Media (2016) LASIK or PRK? Which is right for me? Animation. Filmed [Jan 2016]. YouTube video, 03:50, Posted [Jan 2016]. https://www.youtube.com/watch?v=dKANhIU7Sxk
6. Sinjab Academy (2015) Concepts in refractive surgery—part 2. Filmed [Oct 2015]. YouTube video, 22:10, Posted [Oct 2015]. https://www.youtube.com/watch?v=E2unBK5FkLI
7. Mazzotta C, Raiskup F, Baiocchi S, Scarcelli G, Friedman MD, Traversi C (2017) ACXL beyond Keratoconus: post-LASIK ectasia, post-RK ectasia and pellucid marginal degeneration. In: Management of early progressive corneal ectasia. Springer, Cham, pp 169–196
8. Paul T, Lim M, Starr CE, Lloyd HO, Jackson Coleman D, Silverman RH (2008) Central corneal thickness measured by the Orbscan II system, contact ultrasound pachymetry, and the Artemis 2 system. J Cataract Refract Surg 34(11):1906–1912
9. Liu Z, Huang AJ, Pflugfelder SC (1999) Evaluation of corneal thickness and topography in normal eyes using the Orbscan corneal topography system. Br J Ophthalmol 83(7):774–778
10. Matsuda J, Hieda O, Kinoshita S (2008) Comparison of central corneal thickness measurements by Orbscan II and Pentacam after corneal refractive surgery. Jpn J Ophthalmol 52(4):245

11. Colling AJ (2010) A comparison of three methods of measuring central corneal thickness in normal and thinned corneas. Ph.D. dissertation, The Ohio State University
12. Oliveira Cristina M, Ribeiro Celina, Franco Sandra (2011) Corneal imaging with slit-scanning and Scheimpflug imaging techniques. Clin Exp Optom 94(1):33–42
13. Huang D, Swanson EA, Lin CP, Schuman JS, Stinson WG, Chang W, Hee MR, Flotte T, Gregory K, Puliafito CA (1991) Optical coherence tomography. Science 254(5035):1178–1181
14. Fujimoto JG, Drexler W, Schuman JS, Hitzenberger CK (2009) Optical coherence tomography (OCT) in ophthalmology: introduction. Opt Express 17(5):3978–3979
15. Hurmeric V, Yoo SH, Mutlu FM (2012) Optical coherence tomography in cornea and refractive surgery. Expert Rev Ophthalmol 7(3):241–250
16. Otchere H, Sorbara L (2017) Repeatability of topographic corneal thickness in keratoconus comparing Visante™ OCT and Oculus Pentacam HR® topographer. Contact Lens Anter Eye 40(4):217–223
17. Martínez-Albert N, Esteve-Taboada JJ, Montés-Micó R (2018) Repeatability assessment of anterior segment biometric measurements under accommodative and nonaccommodative conditions using an anterior segment OCT. Graefe's Archive Clin Exp Ophthalmol 256(1):113–123
18. Ramesh PV, Jha KN, Srikanth K (2017) Comparison of central corneal thickness using anterior segment optical coherence tomography versus ultrasound pachymetry. J Clin Diagn Res JCDR 11(8):NC08
19. Antonios R, Abdul Fattah M, Maalouf F, Abiad B, Awwad ST (2016) Central corneal thickness after cross-linking using high-definition optical coherence tomography, ultrasound, and dual scheimpflug tomography: a comparative study over one year. Am J Ophthalmol 167:38–47
20. Xu Z, Peng M, Jiang J, Yang C, Zhu W, Fan L, Shen M (2016) Reliability of Pentacam HR thickness maps of the entire cornea in normal, post-laser in situ Keratomileusis, and keratoconus eyes. Am J Ophthalmol 162:74–82
21. Temstet C, Sandali O, Bouheraoua N, Hamiche T, Galan A, El Sanharawi M, Basli E, Laroche L, Borderie V (2015) Corneal epithelial thickness mapping using Fourier-domain optical coherence tomography for detection of form fruste keratoconus. J Cataract Refract Surg 41(4):812–820
22. Kuerten D, Plange N, Koch EC, Koutsonas A, Walter P, Fuest M (2015) Central corneal thickness determination in corneal edema using ultrasound pachymetry, a Scheimpflug camera, and anterior segment OCT. Graefe's Archive Clin Exp Ophthalmol 253(7):1105–1109
23. Randleman JB, Lynn MJ, Perez-Straziota CE, Weissman HM, Kim SW (2015) Comparison of central and peripheral corneal thickness measurements with scanning-slit, Scheimpflug and Fourier-domain ocular coherence tomography. Br J Ophthalmol 99(9):1176–1181
24. Lin C-W, Wang T-H, Huang Y-H, Huang J-Y (2013) Agreement and repeatability of central corneal thickness measurements made by ultrasound pachymetry and anterior segment optical coherence tomography. Taiwan J Ophthalmol 3(3):98–102
25. Yazici AT, Bozkurt E, Alagoz C, Alagoz N, Pekel G, Kaya V, Yilmaz OF (2010) Central corneal thickness, anterior chamber depth, and pupil diameter measurements using Visante OCT, Orbscan, and Pentacam. J Refract Surg 26(2):127–133
26. Doors M, Cruysberg LPJ, Berendschot TTJM, de Brabander J, Verbakel F, Webers CAB, Nuijts RMMA (2009) Comparison of central corneal thickness and anterior chamber depth measurements using three imaging technologies in normal eyes and after phakic intraocular lens implantation. Graefe's Archive Clin Exp Ophthalmol 247(8):1139–1146
27. Cheng ACK, Rao SK, Lau S, Lam DSC, Leung CKS (2008) Central corneal thickness measurements by ultrasound, Orbscan II, and Visante OCT after LASIK for myopia. J Refract Surg 24(4):361–365
28. Kim HY, Budenz DL, Lee PS, Feuer WJ, Barton Keith (2008) Comparison of central corneal thickness using anterior segment optical coherence tomography vs ultrasound pachymetry. Am J Ophthalmol 145(2):228–232
29. Li EYM, Mohamed S, Leung CKS, Rao SK, Cheng ACK, Cheung CYL, Lam DSC (2007) Agreement among 3 methods to measure corneal thickness: ultrasound pachymetry, Orbscan II, and Visante anterior segment optical coherence tomography. Ophthalmology 114(10):1842–1847

30. Ho T, Cheng ACK, Rao SK, Lau S, Leung CKS, Lam DSC (2007) Central corneal thickness measurements using Orbscan II, Visante, ultrasound, and Pentacam pachymetry after laser in situ keratomileusis for myopia. J Cataract Refract Surg 33(7):1177–1182
31. Zhao PS, Wong TY, Wong W-L, Saw S-M, Aung T (2007) Comparison of central corneal thickness measurements by visante anterior segment optical coherence tomography with ultrasound pachymetry. A J Ophthalmol 143(6):1047–1049
32. Haque S, Simpson T, Jones L (2006) Corneal and epithelial thickness in keratoconus: a comparison of ultrasonic pachymetry, Orbscan II, and optical coherence tomography. J Refract Surg 22(5):486–493
33. Leung DYL, Lam DKT, Yeung BYM, Lam DSC (2006) Comparison between central corneal thickness measurements by ultrasound pachymetry and optical coherence tomography. Clin Exp Ophthalmol 34(8):751–754
34. Li Y, Shekhar R, Huang D (2006) Corneal pachymetry mapping with high-speed optical coherence tomography. Ophthalmology 113(5):792–799
35. Lackner B, Schmidinger G, Pieh S, Funovics MA, Skorpik C (2005) Repeatability and reproducibility of central corneal thickness measurement with Pentacam, Orbscan, and ultrasound. Optom Vis Sci 82(10):892–899
36. Fishman GR, Pons ME, Seedor JA, Liebmann JM, Ritch R (2005) Assessment of central corneal thickness using optical coherence tomography. J Cataract Refract Surg 31(4):707–711
37. Wong ACM, Wong CC, Yuen NSY, Hui SP (2002) Correlational study of central corneal thickness measurements on Hong Kong Chinese using optical coherence tomography, Orbscan and ultrasound pachymetry. Eye 16(6):715

Hybrid SIFT Feature Extraction Approach for Indian Sign Language Recognition System Based on CNN

Abhishek Dudhal, Heramb Mathkar, Abhishek Jain, Omkar Kadam and Mahesh Shirole

Abstract Indian sign language (ISL) is one of the most used sign languages in the Indian subcontinent. This research aims at developing a simple Indian sign language recognition system based on convolutional neural network (CNN). The proposed system needs webcam and laptop and hence can be used anywhere. CNN is used for image classification. Scale invariant feature transformation (SIFT) is hybridized with adaptive thresholding and Gaussian blur image smoothing for feature extraction. Due to unavailability of ISL dataset, a dataset of 5000 images, 100 images each for 50 gestures, has been created. The system is implemented and tested using python-based library Keras. The proposed CNN with hybrid SIFT implementation achieves 92.78% accuracy, whereas the accuracy of 91.84% was achieved for CNN with adaptive thresholding.

1 Introduction

Sign language is an efficient and natural way of communication for the hearing-impaired and verbally impaired community. Around 2.7 million of India's total population is hearing-impaired and 98% of this population uses Indian sign language as the primary language for communication. However, hearing-impaired community

A. Dudhal (✉) · H. Mathkar · A. Jain · O. Kadam · M. Shirole
Department of Computer Engineering and Information Technology, Veermata Jijabai Technological Institute, Mumbai, India
e-mail: addudhal_b14@it.vjti.ac.in

H. Mathkar
e-mail: hkmathkar_b14@it.vjti.ac.in

A. Jain
e-mail: amjain_b14@it.vjti.ac.in

O. Kadam
e-mail: oskadam_b14@it.vjti.ac.in

M. Shirole
e-mail: mrshirole@it.vjti.ac.in

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_72

experiences difficulties while communicating with people who lack knowledge of sign language. A human translator is required to translate sign language into speech. This solution is translator dependent and it fails in absence of a translator. A differently abled person can be empowered with a computer-based system for translation. A computer-based system can be trained to recognize ISL efficiently, thereby providing high availability, ease of use, and efficient navigation and trade to differently abled persons.

For Indian sign language recognition, the research work undertaken ranges from introducing a smart glove to monitor movements of fingers and hand to image processing that analyzes hand gestures captured. Heera et al. [1] introduced sensors incorporated glove-based approach to convert ISL into speech with the help of Bluetooth module and an Android smartphone. Ekbote et al. [2] proposed a method for ISL recognition using artificial neural network (ANN) [3] and support vector machine (SVM) [4] classifiers. Authors have used a self-created dataset for ISL, 0–9 numbers, which is very limited. Histogram of oriented gradients (HOG) [5] and ANN-based approach proposed by them was able to achieve 99% accuracy. Beena et al. [6] proposed a CNN [7] based ASL recognition system. They used ASL dataset with 33,000 images for 24 alphabets and 0–9 numbers and accuracy of 94.6774% was achieved. Pigou et al. [8] were able to classify 20 Italian gestures using CNN with validation accuracy of 91.7%. They used Microsoft Kinect to capture gestures. Microsoft Kinect is able to capture depth feature. Depth feature aids significantly in image classification.

The contemporary research focused on the numbers, alphabets, limited words, and single-handed gestures. In contrast, this paper aims to help the hearing-impaired community by developing a simple computer vision-based system, which works on 50 ISL words including numbers and double handed gestures.

This paper is composed of six sections; Sect. 2 discusses the basic concepts used in the paper. Section 3 discusses the proposed system. Results of the experiment are discussed in Sect. 4. A complete conclusion is drawn in Sect. 5. Section 6 highlights the future aspects of the paper.

2 Basic Concepts

2.1 Adaptive Thresholding

Image binarization can be achieved with the help of adaptive thresholding. Image binarization is a method of separation of pixel intensity in two groups. Setting black as foreground and white as background or vice versa. Image binarization and thresholding is an effective way to separate an object from the background. In adaptive thresholding, a threshold value is set such that pixel intensity below that threshold will be treated as zero while pixel intensity greater than the threshold will be treated as one. Equation 1 shows the formula for adaptive thresholding.

$$b(x, y) = \begin{cases} 0, \ I(x, y) \leq T(x, y) \\ 1, \ I(x, y) > T(x, y) \end{cases} \tag{1}$$

where $T(x, y)$ is the threshold, $b(x, y)$ is the binarized image, and $I(x, y)$ is the intensity of pixel at (x, y).

2.2 Image Smoothing Using Gaussian Blur

Image smoothing is a technique of removing noise from digital images. Smoothing can remove noise without losing important features from the image. Gaussian blur filter [9] uses a Gaussian function [10] to calculate transformation. Equation 2 represents Gaussian blur operator.

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{\frac{-(x^2+y^2)}{2\sigma^2}} \tag{2}$$

2.3 Key-Point Generation Using Scale Invariant Feature Transform

SIFT [11] is scale, rotation, viewpoint, illumination invariant algorithm. The key-point generation involves three steps. The first step is the generation of scale space. In the second step, Laplacian of Gaussians (LoG) [12] is generated while in the final step, key points are calculated.

Scale-Space Generation In the scale-space generation step, the original image is taken, and progressively blurred out images are generated. Then, the original image is resized to half and blurred out images are generated again. Images of the same size form an octave. The number of scales and octaves depend on the user. Blurring can be thought of as a convolution of the Gaussian operator and the image.

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y) \tag{3}$$

where L is the obtained Blurred image, G is the Gaussian blur operator, I is the image, (x, y) are the coordinates in the image, σ is scale parameter, and * is convolution operator in (x, y). This operator applies Gaussian blur G onto the image I. Gaussian blur function is represented by Eq. 2. Let the amount of the blurring in one image be σ and then, the amount of blurring in the next image will be $k\sigma$, where k is a constant chosen by the user.

Laplacian of Gaussian (LoG) Approximation In the generation of LoG [12] step, an image is taken and blurred a little using Gaussian blur. Then, second-order derivatives are calculated on it (Laplacian). This process is computationally expensive. So to calculate LoG quickly, scale space obtained in the previous step is used. Difference between two consecutive scales is calculated. These differences of Gaussian images are approximately equal to LoG.

Finding Key Points In this step, each and every pixel is iterated, and all of its neighbors are checked. The check is done within the current image, one above and one below it. The pixel is marked as an approximate key point if it is greatest or least of all 26 neighbors. The minima or maxima lies somewhere between the pixels. So, subpixel value is calculated mathematically using Taylor expansion [13] of the image around the approximate key point.

2.4 CNN

CNN is a deep learning neural network. CNN is specialized for images, audios, videos, and speech processing. It is designed to learn features with the help of filters and hence requires very little data preprocessing and feature extraction. CNN is composed of one input layer, one or multiple convolutional layers, one or multiple max-pooling layers, a fully connected layer, and an output layer. Input layer accepts input which is passed to next layer. The convolutional layer is responsible to apply convolution operation to the input data. This layer works as eyes of CNN and looks for specific features useful for classification. The Filters are also known as the kernels. Max-pooling layer is useful for reducing parameters size and hence processing time. Fully connected layer acts as a classifier. Output layer gives an output vector consisting of probability for different classes. Each neuron uses activation function for mathematical processing of data. CNN has mechanism of dropout which is used to avoid overfitting.

2.5 Confusion Matrix

The confusion matrix is a matrix which is used to summarize the performance of the classifier. For a good classifier, it is a sparse matrix and can be represented in the form of a graph. Actual class is represented on X-axis while predicted class is represented on the Y-axis. Label to point (X, Y) represents a number of the example for which actual class is X and predicted is Y. When $X = Y$, then it is treated as accurate classification. Hence, (X, X) are treated as correctly classified examples. The confusion matrix is used to analyze the results of CNN with hybrid SIFT implementation, later in this research.

Precision Precision is a fraction of correctly identified examples to the number of examples for which that particular class is predicted as positive. Equation 4 specifies the formula for precision in the multiclass classifier.

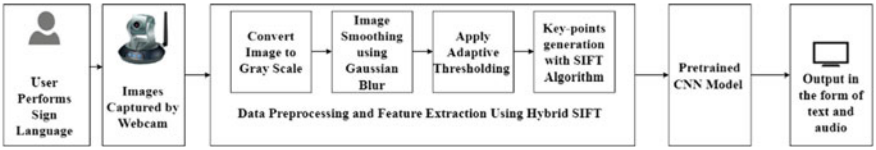


Fig. 1 Proposed system for ISL classification

$$\text{Precision}_i = \frac{\text{CF}_{ii}}{\sum_1^n \text{CF}_{ij}} \tag{4}$$

where CF_{ii} is (i, i)th entry in confusion matrix, CF_{ij} is (i, j)th entry in confusion matrix, and n is the total number of classes.

Recall A recall is a fraction of correctly identified examples to the number of examples available for that class. Equation 5 specifies the formula for recall in a multiclass classifier.

$$\text{Recall}_i = \frac{\text{CF}_{ii}}{\sum_1^n \text{CF}_{ji}} \tag{5}$$

where CF_{ii} is (i, i)th entry in confusion matrix, CF_{ji} is (j, i)th entry in confusion matrix, and n is total number of classes.

3 Proposed System

Figure 1 shows a flow diagram of a proposed system for ISL classification. Gestures performed by the user are captured by a webcam. The captured gesture is preprocessed and features are extracted using hybrid SIFT. Hybrid SIFT is discussed in Sect. 3.1. The preprocessed gesture is fed to a pretrained CNN model. The CNN model used for this research is explained in Sect. 3.2.

3.1 Data Preprocessing and Feature Extraction

Though CNN has the ability to work without any feature extraction and data preprocessing, data preprocessing is used for reducing computational power required and for better performance of the model. This paper presents a hybrid SIFT approach for data preprocessing and feature extraction.

Hybrid SIFT As discussed in Sect. 2.3, SIFT calculates key points. Data preprocessing before applying SIFT on the image can reduce noise and can help in better key-points generation. The key-point generation step of the SIFT is hybridized with

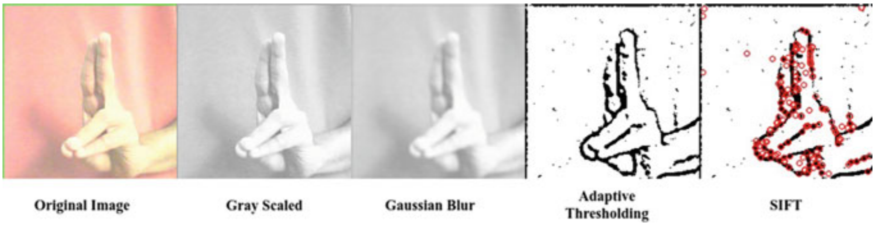


Fig. 2 Application of SIFT on preprocessed image (hybrid SIFT)

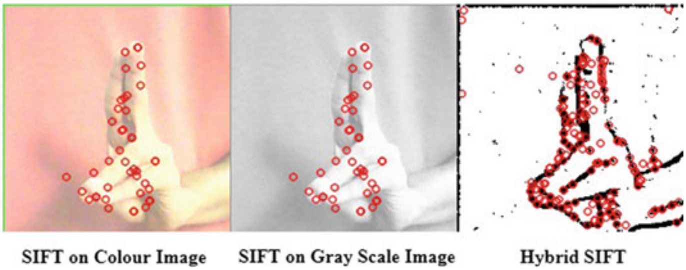


Fig. 3 Key-point calculation on original image with SIFT, grayscaled image with SIFT and hybrid SIFT

Gaussian blurring and adaptive thresholding. Steps involved in performing hybrid SIFT are shown in Fig. 2. In the first step, the image is captured and resized to 200 × 200 pixels. The second step involves converting the resized original image to grayscaled image. In the third step as discussed in Sect. 2.2, the grayscaled image is smoothened using a Gaussian blur filter. In next step, the smoothened grayscaled image is binarized using adaptive thresholding. Finally, key-point generation step of SIFT algorithm is applied to the binarized image. Figure 3 shows a comparison of key-point calculation using SIFT algorithm on a simple grayscaled image and hybrid SIFT algorithm. Figure 3 suggests that SIFT applied on the adaptive thresholded image gives better key points than when applied directly to the original image.

3.2 Architecture of CNN Model

This paper presents a self-designed CNN model for gesture recognition. Figure 4 shows the code snippet of the proposed CNN model. Proposed CNN model has 10 convolutional layers. Convolutional layers are divided into five groups. Each group contains two convolutional layers. For the first group, the number of filters is kept 32, for second group 64, for third group 128, for fourth group 256, and for the final group, it is kept 512. Each convolutional layer has a filter size of 3 × 3. Max-pooling

```
model = Sequential()
model.add(Conv2D(32, (3, 3), padding='valid', activation='relu', input_shape=input_shape))
model.add(Conv2D(32, (3, 3), activation='relu'))
model.add(MaxPooling2D(pool_size=(2, 2)))
model.add(Dropout(0.25))
model.add(Conv2D(64, (3, 3), padding='valid', activation='relu'))
model.add(Conv2D(64, (3, 3), activation='relu'))
model.add(MaxPooling2D(pool_size=(2, 2)))
model.add(Dropout(0.25))
model.add(Conv2D(128, (3, 3), padding='valid', activation='relu'))
model.add(Conv2D(128, (3, 3), activation='relu'))
model.add(MaxPooling2D(pool_size=(2, 2)))
model.add(Dropout(0.25))
model.add(Conv2D(256, (3, 3), padding='valid', activation='relu'))
model.add(Conv2D(256, (3, 3), activation='relu'))
model.add(MaxPooling2D(pool_size=(2, 2)))
model.add(Dropout(0.25))
model.add(Conv2D(512, (3, 3), padding='valid', activation='relu'))
model.add(Conv2D(512, (3, 3), activation='relu'))
model.add(MaxPooling2D(pool_size=(2, 2)))
model.add(Dropout(0.25))
model.add(Flatten())
model.add(Dense(512, activation='relu'))
model.add(Dropout(0.25))
model.add(Dense(no_classes, activation='softmax'))
model.compile(loss='categorical_crossentropy', optimizer='adadelta', metrics=['accuracy'])
```

Fig. 4 Code snippet for CNN model

layer with pooling window of 2×2 is applied after each group of the convolutional layer. After each max-pooling layer, to avoid overfitting dropout ratio of 0.25 is used. Each convolution layer uses rectified linear unit (RELU) as the activation function [14]. Optimizer for training was kept fixed as Adadelta [15]. The model uses Softmax [16] as a classifier. To calculate the performance of model, the loss function is used as cross-entropy [17]. The CNN model is coded by using python-based library Keras [18].

4 Results

The proposed CNN model is applied to self-created dataset of ISL with 50 signs and 100 images per sign. CNN model is trained with an incremental increase of epochs approach. Batch size for the model training is kept fixed at 64. Starting from 10 epochs, epochs were increased by 5 till model converged with constant validation accuracy. With increments in the number of epoch approach, proposed model converged with constant validation accuracy at 25 epochs. The model was trained till 50 epochs. Section 4.1 describes the system requirement for the proposed system, Sect. 4.2 elaborates data acquisition process. The accuracy of the proposed CNN model is discussed in Sect. 4.3 which shows a confusion matrix generated for the proposed model and dataset. Finally, we discuss the comparison of the proposed system with related research.

4.1 System Requirements

The proposed system tried to keep user interaction with the system using desktop application. System requirements are as follows:

- 4 GB Ram
- 1 GB Free Space
- Web Cam.

4.2 Data Acquisition

The standard dataset for Indian sign language is not available. Two sets of datasets each of 5000 images were created. One was used for training and validation purpose while other for the testing purpose. Dataset was created with the help of 5 MP webcam attached to a laptop. Dictionary of 50 most used signs is created by taking help of Deaf and Dumb School, Mumbai. Each dataset contains total 5000 images, where for each sign 100 images of 200×200 pixel are captured. Dataset is created with the help of 20 people. The age group varies from 16 to 50. Around 70% of the people were in the age group of 20–25. Both males and females have participated in dataset creation.

4.3 Accuracy of Proposed CNN Model

Table 1 summarizes accuracy matrix of the proposed model. For CNN with adaptive thresholding, training accuracy (TA) achieved is 97.58% and validation accuracy (VA) is 91.84%. For CNN with hybrid SIFT, training accuracy increases to 98.83% and validation accuracy increases to 92.72%. It shows that CNN with hybrid SIFT performs better than CNN with only adaptive thresholding. Figure 5 shows the accuracy graph of the proposed model. In the accuracy graph, accuracy is represented as dependent variable on Y-axis and number of epochs as an independent variable on X-axis. Accuracy graph lists accuracy for training dataset by blue line and accuracy on validation dataset by the orange line. Accuracy graph shown in Fig. 5 suggests that early stopping of the model at around 25 epoch can avoid overfitting issue.

Table 1 Accuracy matrix for proposed CNN model

CNN with adaptive thresholding		CNN with hybrid SIFT	
TA	VA	TA	VA
97.58	91.84	98.83	92.72

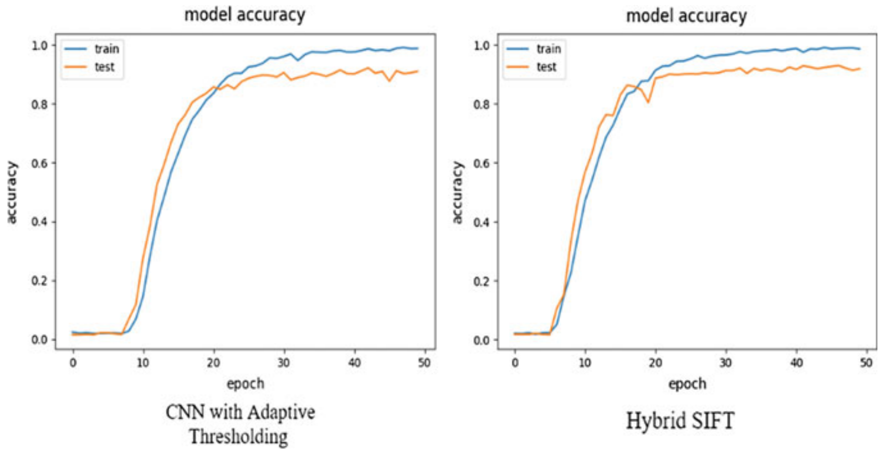


Fig. 5 Model accuracy

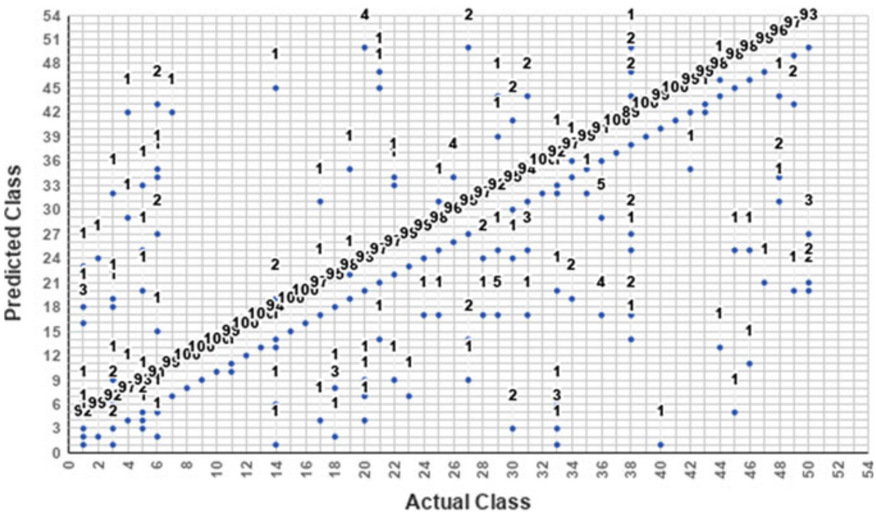


Fig. 6 Confusion matrix obtained by running trained CNN model on testing dataset

4.4 Confusion Matrix of Proposed System

The hybrid SIFT algorithm is first applied on the dataset and then CNN model is trained by using the same dataset. As discussed in Sect. 2.5, the confusion matrix is created by running trained CNN model on the testing dataset. Figure 6 shows the confusion matrix in a graphical format. The *Y*-axis of the graph represents predicted class while *X*-axis represents the actual class. Confusion matrix obtained for the proposed model is of sparse nature.

Table 2 Precision and recall obtained by running trained CNN model on testing dataset

Sign	Precision	Recall	Sign	Precision	Recall	Sign	Precision	Recall
House	94.8454	92	High	97.9381	95	Rose	97.0588	99
A board	97.0588	99	How many	95.1456	98	See	98.913	91
All Gone	92.9293	92	I	94.898	93	Seven	99.0099	100
Baby	96.0396	97	Man	96.0396	97	Short	100	89
Beside	97.8947	93	Marry	98.9796	97	Six	99.0099	100
Book	91.9192	91	Meat	99	99	Superior	100	99
Bowl	97.0588	99	Medal	96.1165	99	Ten	98.0392	100
Bridge	98.0392	100	Mid Day	92.4528	98	Thick	97.0588	99
Camp	96.1538	100	Middle	100	96	Thin	96.1165	99
Cartridge	99.0099	100	Money	93.1373	95	Three	94.2308	98
Eight	99	99	Moon	100	97	Tobacco	98	98
Five	100	100	Mother	93.8775	92	Two	98.9899	98
Fond	98.0392	100	Nine	100	95	Up	97.0588	99
Four	95.9184	94	One	96.9072	94	Watch	100	96
Friend	99.0099	100	Opposite	97.0874	100	Write	100	97
Glove	97.0874	100	Prisoner	97.8723	92	You	93	93
Hang	86.6071	97	Ring	92.381	97			

Precision and recall were calculated for multiclass using formula shown in Sect. 2.5. Table 2 summarizes precision and recall in percent for proposed model and 50 signs available in the dataset. Highest precision obtained is 100% while least precision is 86.6071%. Highest recall obtained is 100% while least recall is 89%.

5 Conclusion

The sensor incorporated glove-based systems which are most common in sign language recognition are usually costly and difficult to use; in contrast, image classification-based system proposed in this paper is much cheaper and easier to use. CNN is an important and efficient algorithm for image classification. This paper proposed a system for Indian sign language recognition using classification by CNN and feature extraction by hybrid SIFT. CNN is robust and stable such that it requires very little image preprocessing. But image processing using hybrid SIFT improves the performance of CNN classifier. The model proposed in this paper has achieved validation accuracy of 92.78% for CNN with hybrid SIFT approach while 91.84% accuracy was achieved for CNN with adaptive thresholding approach. The system proposed in this paper can work just with a laptop and web camera and hence can be used with mobility by the hearing-impaired community. The system proposed in this paper can also be used for learning Indian sign language.

6 Future Work

Dataset used in this paper contains 50 Indian signs. Further work can be done to increase the number of signs as well as images per sign. This paper considers only static Indian signs. In the future, CNN can also be implemented for motion-based Indian signs. The proposed system can be extended to work with handheld mobile devices by optimizing memory and power requirement.

Acknowledgements We would like to thank Principal of Deaf and Dumb School, Mumbai for her help in understanding ISL gestures. We would like to thank teachers and students of Deaf and Dumb School for their help in the creation of ISL dataset.

References

1. Heera SY et al (2017) Talking hands—an Indian sign language to speech translating gloves. In: 2017 International conference on innovative mechanisms for industry applications (ICIMIA)
2. Ekbote J et al (2017) Indian sign language recognition using ANN and SVM classifiers. In: 2017 International conference on innovations in information, embedded and communication systems (ICIIECS)
3. Duch W (2005) Artificial neural network biological inspirations. In: ICANN 2005: 15th international conference, Warsaw, Poland, September 11–15, 2005, proceedings, Pt. 1. Springer, Heidelberg
4. Qi X et al (2017) Data classification with support vector machine and generalized support vector machine
5. Vo T et al (2015) Tensor decomposition and application in image classification with histogram of oriented gradients. Neurocomputing 165:38–45
6. Beena MV, Agnisarman Namboodiri MN (2017) Automatic sign language finger spelling using convolution neural network: analysis. Int J Pure Appl Math 117(20)
7. Aghdam HH, Heravi EJ (2017) Convolutional neural networks. In: Guide to convolutional neural networks, pp 85–130
8. Pigou L et al (2015) Sign language recognition using convolutional neural networks. Computer vision—ECCV 2014 workshops lecture notes in computer science, pp 572–578
9. Gaussian Smoothing (2008) Wolfram Demonstrations Project
10. Gaussian function. Springer Reference
11. Lowe DG (2004) Distinctive image features from scale-invariant keypoints. Int J Comput Vision 60(2):91–110
12. Laplacian of Gaussian Filtering (2008) Wolfram Demonstrations Project
13. Berz M (1997) From Taylor series to Taylor models
14. Ramachandran P et al (2017) Searching for activation functions. In: ICLR 2018 conference

15. Zeiler MD (2012) ADADELTA: an adaptive learning rate method
16. Pellegrini T (2015) Comparing SVM, Softmax, and shallow neural networks for eating condition classification. Interspeech 2015
17. Dahal P, Classification and loss evaluation—Softmax and cross entropy loss. https://deepnotes.io/softmax-crossentropy
18. Keras: The Python Deep Learning library. https://keras.io/

A Contemporary Framework and Novel Method for Segmentation of Brain MRI

A. Jagan

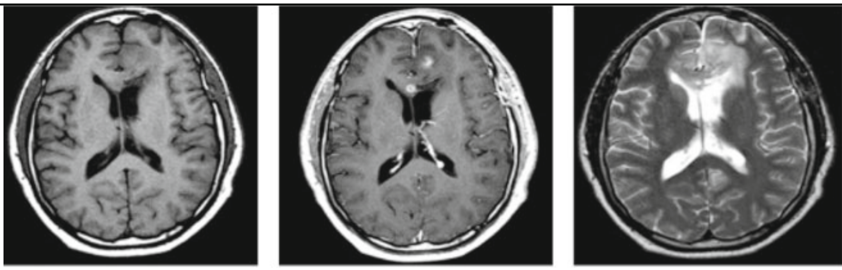
Abstract Brain magnetic resonance imaging plays a vital role in medical image processing for detection of brain tumor, therapy response evaluation, brain tumor diagnosis, and treatment selection. Contrast-enhanced T1C 3D brain magnetic resonance imaging is extensively used brain-imaging modalities. Automated segmentation and detection of brain tumors in contrast-enhanced T1C 3D brain magnetic resonance imaging is a very complicated task due to high disparity in shape, size, and appearance of brain tumor. Nevertheless, so many analyses have been conducted in the similar research work, it remains a very complicated task for automated segmentation and detection of brain tumor in magnetic resonance imaging, and enhancing segmentation accuracy of brain tumor is still continuing field. The main aim of this research work is to develop a fully automated segmentation framework for segmentation and detection of the brain tumor that is allied with contrast-enhanced T1C 3D brain magnetic resonance imaging. Consequently, this research work deals about development of segmentation framework for detection of tumor in brain 3D MR images. The proposed segmentation framework ingrates the most established fuzzy C means clustering method and improved Expectation Maximization (EM) method. An anisotropic filter is used for preprocessing and subsequently, it is employed to the fuzzy C means clustering method, improved Expectation Maximization (EM), and proposed method for superior segmentation and detection of tumor. The performance result of proposed framework is evaluated on 10 patients' simulated medical clinical dataset. The performance results of the proposed framework exhibited better results as compared with presented methods.

1 Introduction

The advancement in the segmentation and detection of brain tumor in magnetic resonance imaging is evolving. The similar research has exposed the various methods

A. Jagan (✉)
CSE Department, BVRIT, Narsapur, India
e-mail: jagan.amgoth@bvrit.ac.in

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_73



T1 Brain MRI	T1C Brain MRI	T2 Brain MRI
T1 Brain Magnetic Resonance Imaging that is often used to illustrate the fat region in the Brain.	T1C Brain Magnetic Resonance Imaging highlights the tumor regions in the Brain.	T2 Brain Magnetic Resonance Imaging that is used to represent the water region in the Brain.

Fig. 1 Magnetic resonance imaging series

and potential techniques to the brain image analysis and detection of tumor in 3D brain magnetic resonance imaging. The MRI images present the noninvasive and broad range of visualization of brain internal anatomical structure and analysis of the brain. The major research direction is focused on development of the segmentation accuracy for detection of tumor in brain magnetic resonance imaging. Hence, this research work deals about the analysis of contrast-enhanced T1C 3D brain magnetic resonance imaging with the intricate and superior quality became the most challenge job for magnetic resonance imaging analysis experts [1, 2]. Furthermore, because of the human intervention, the analysis of brain magnetic resonance imaging is bound to be invalid. Further, magnetic resonance imaging by manual analyses takes lots of time and very limited to discriminate of tumor in contrast-enhanced T1C 3D brain magnetic resonance imaging as compared with the modern automatic methods [1, 3, 4].

The usually used technique for analysis of brain magnetic resonance imaging is segmentation method that is most popular for imaging of medical applications especially brain magnetic resonance imaging. The limitations recognized in the research study exposes for assorted segmentation methods are inadequate in generating improved segmentation accuracy and more often than not focused on segmentation and detection of the brain tumor. The present research work focused on the incomparable segmentation accuracy results [4–6]. The brain magnetic resonance imaging has diverse MRI series such as T1 brain magnetic resonance imaging, contrast-enhanced T1C 3D brain MR imaging, and T2 brain magnetic resonance imaging as shown in Fig. 1.

This paper is focused on the segmentation of contrast-enhanced T1C 3D brain magnetic resonance imaging. The usefulness of T1C series has been estimated in various ways in the central nervous system, i.e., infarction, head injuries, subarachnoid hemorrhage, and multiple sclerosis. Contrast-enhanced T1C brain magnetic resonance imaging provides superior anatomical details had better than other brain MR imaging. In preprocessing stage, the anisotropic filtering [7, 8] is used to smooth edges of brain magnetic resonance imaging while preserving edges and it also used to improve the quality of input magnetic resonance imaging without losing the boundaries. The applications of anisotropic filter have fully extended, and it is now used for denoising of medical images and medical image enhancement, etc. The anisotropic filter is focused on input magnetic resonance imaging variance and standard deviation [7]. Fuzzy C means clustering method [9–13]. In this work, we analyze the accuracy of well-known methods, i.e., improved Expectation Maximization (EM) [5, 14–16] and Proposed method for segmentation and detection of brain disorder in contrast-enhanced T1C 3D brain magnetic resonance imaging. Hereafter the rest of research work is organized to obtain better segmentation accuracy and detection of tumor in contrast-enhanced T1C 3D brain magnetic resonance imaging.

The rest of the research work is organized as follows: Sect. 2 in this research work presents the Proposed Methodology and approach for enhancement of segmentation results, in Sect. 3 discuss in detail about applications and the evaluation of results which is executed on multiple contrast-enhanced T1C magnetic resonance imaging datasets and in Sect. 4 discuss the conclusions of the research work that is presented in this paper.

2 Proposed Methodology

The major objective of this proposed research work is to develop an automated segmentation framework for the detection of brain tumor in contrast-enhanced T1C 3D brain magnetic resonance imaging. In this research work, initially, the entire brain imaging is segmented with fuzzy C means clustering method and the improved Expectation Maximization (EM). These segmentation results are optimally integrated with proposed method to improve the segmentation accuracy. The main idea behind this work is to integrate the most established segmentation methods, i.e., fuzzy C means clustering method and the improved Expectation Maximization (EM). Consequently, the proposed framework is equipped with the anisotropic filter in the preprocessing stage to improve the quality of input brain MR image and for the better segmentation and detection of brain tumor. The block diagram of the proposed research work is shown in Fig. 2.

(a) FCMC Method = {CONTRAST-ENHANCED T1C $3D_1$_FCMCSeg, CONTRAST-ENHANCED T1C $3D_2$_FCMCSeg, CONTRAST-ENHANCED T1C $3D_3$_FCMCSeg, …, CONTRAST-ENHANCED T1C $3D_n$_FCMCSeg}.
(b) EM Method = {CONTRAST-ENHANCED T1C $3D_1$_EMSeg, CONTRAST-ENHANCED T1C $3D_2$_EMSeg, CONTRAST-ENHANCED T1C $3D_3$_EMSeg, …, CONTRAST-ENHANCED T1C $3D_n$_EMSeg}.

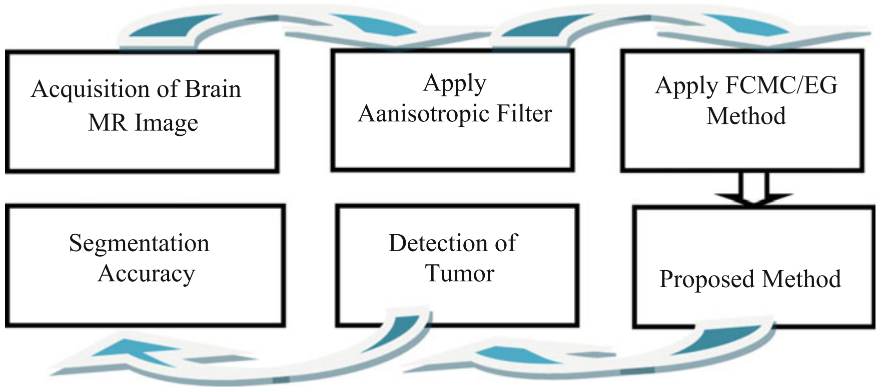


Fig. 2 Block diagram of the proposed work

(c) Proposed Method = {CONTRAST-ENHANCED T1C $3D_1$_AMCSeg, CONTRAST-ENHANCED T1C $3D_2$_AMCSeg, CONTRAST-ENHANCED T1C $3D_3$_AMCSeg, …, CONTRAST-ENHANCED T1C $3D_n$_AMCSeg}.

(d) Segmentation Results (Tumor Detection) of the proposed Method = {$Tumor_1$_SegImage, $Tumor_2$_SegImage, $Tumor_3$_SegImage, …, $Tumor_n$_SegImage}.

(e) Ground truth(Images) Dataset = {CONTRAST-ENHANCED T1C $3D_1$_Truth, CONTRAST-ENHANCED T1C $3D_2$_Truth, CONTRAST-ENHANCED T1C $3D_3$_Truth, …, CONTRAST-ENHANCED T1C $3D_n$_Truth}.

The proposed procedure for automated segmentation for detection of tumor in contrast-enhanced T1C 3D brain magnetic resonance imaging:

Input: MRI Dataset – {CONTRAST-ENHANCED T1C $3D_1$, CONTRAST-ENHANCED T1C $3D_2$, CONTRAST-ENHANCED T1C $3D_3$, …, CONTRAST-ENHANCED T1C $3D_n$}.

Output: Brain tumor detection in Contrast-Enhanced T1C 3D MR Image – {$Tumor_1$_SegImage, $Tumor_2$_SegImage, $Tumor_3$_SegImage, …, $Tumor_n$_SegImage}.

Step 1 Preprocessing: The Anisotropic Filter (AF) is applied on input Original input MRI Dataset {CONTRAST-ENHANCED T1C $3D_1$, CONTRAST-ENHANCED T1C $3D_2$, CONTRAST-ENHANCED T1C $3D_3$, …, CONTRAST-ENHANCED T1C $3D_n$} to enhance and improve the quality input MRI.

Step 2 Initial Segmentation: Primarily Improved Expectation Maximization (EM) method and fuzzy C means clustering method were functional on preprocessed brain MR images which is {CONTRAST-ENHANCED T1C $3D_1$_PreP, CONTRAST-ENHANCED T1C $3D_2$_PreP, CONTRAST-ENHANCED T1C $3D_3$_PreP, …, CONTRAST-ENHANCED T1C $3D_n$_PreP}.

a. Segmented results of FCMC is denoted as {CONTRAST-ENHANCED T1C $3D_1$_FCMCSeg, CONTRAST-ENHANCED T1C $3D_2$_FCMCSeg, CONTRAST-ENHANCED T1C $3D_3$_FCMCSeg, …, CONTRAST-ENHANCED T1C $3D_n$_FCMCSeg}.
b. Segmented results of EM method is denoted as {CONTRAST-ENHANCED T1C $3D_1$_EMSeg, CONTRAST-ENHANCED T1C $3D_2$_EMSeg, CONTRAST-ENHANCED T1C $3D_3$_EMSeg, …, CONTRAST-ENHANCED T1C $3D_n$_EMSeg}.

Step 3 Proposed Integration Method is optimally integrating the segmentation results of Step 2 (a) and (b).
The outcome of optimal integration is denoted as {CONTRAST-ENHANCED T1C $3D_1$_AMCSeg, CONTRAST-ENHANCED T1C $3D_2$_AMCSeg, CONTRAST-ENHANCED T1C $3D_3$_AMCSeg, …, CONTRAST-ENHANCED T1C $3D_n$_AMCSeg}.

Step 4 Detection of the Tumor: The proposed method is denoted as {CONTRAST-ENHANCED T1C $3D_1$_AMCSeg, CONTRAST-ENHANCED T1C $3D_2$_AMCSeg, CONTRAST-ENHANCED T1C $3D_3$_AMCSeg, …, CONTRAST-ENHANCED T1C $3D_n$_AMCSeg}.

Step 5 The segmentation accuracy: Segmented contrast-enhanced T1C 3D MR image is compared with corresponding ground truth image for the performance evaluation of the proposed method, and it is evaluated over of improved Expectation Maximization (EM), fuzzy C means clustering method by using standard performance metrics, i.e., Segmentation Accuracy.

3 Results and Discussions

In this research work presents the MATLAB implementation of the proposed framework that is evaluated the performance of the current improved Expectation Maximization (EM) method, fuzzy C means clustering method, and proposed method on contrast-enhanced T1C 3D Brain MR Images as shown in Fig. 3.

The relative performance of the presented methods, i.e., fuzzy C means clustering method, improved Expectation Maximization (EM) method and proposed method have been demonstrated on BRATS simulated contrast-enhanced T1C 3D Brain MR Image datasets. The BRATS datasets that hold the simulated 3D brain MR images with corresponding ground truth image for analysis of segmentation accuracy as shown in Fig. 3. The visual segmentation result to the detection of tumor in contrast-enhanced T1C 3D Brain MR Image using improved Expectation Maximization (EM) method, fuzzy C means clustering method, and proposed method as shown in Table 1 and Fig. 3.

The relative performance of the presented methods, i.e., fuzzy C means clustering method, improved Expectation Maximization (EM) method, and proposed method

Fig. 3 T1C brain magnetic resonance imaging with segmented tumor regions

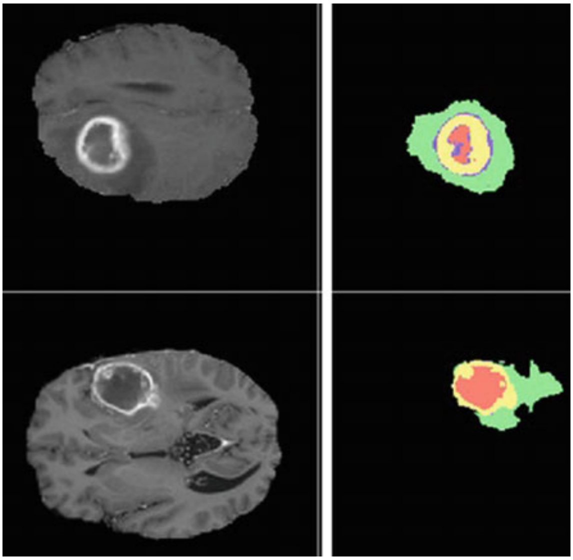


Table 1 Analysis of segmentation accuracy

Input 3D MR image	EG method (%)	FCMC method (%)	Proposed method (%)
Patient-1	92.31	96.22	97.27
Patient-2	96.81	96.56	97.29
Patient-3	93.31	96.91	97.7
Patient-4	93.31	95.22	96.27
Patient-5	97.11	95.56	97.79
Patient-6	94.38	95.31	96.71
Patient-7	98.31	97.13	98.91
Patient-8	97.19	97.21	97.48
Patient-9	96.31	98.13	98.78
Patient-10	96.99	96.21	97.78

have been demonstrated on BRATS simulated contrast-enhanced T1C 3D Brain MR Image datasets as shown in Fig. 4.

The proposed method achieved 97.98% of the average segmentation accuracy on 10 patients' contrast-enhanced T1C 3D Brain MR Image dataset that is superior compared with current methods and Comparison of average segmentation accuracy as shown in Fig. 5.

The research work clearly demonstrates the improved performance of proposed method, and it is compared to the current method on contrast-enhanced T1C 3D brain magnetic resonance imaging dataset.

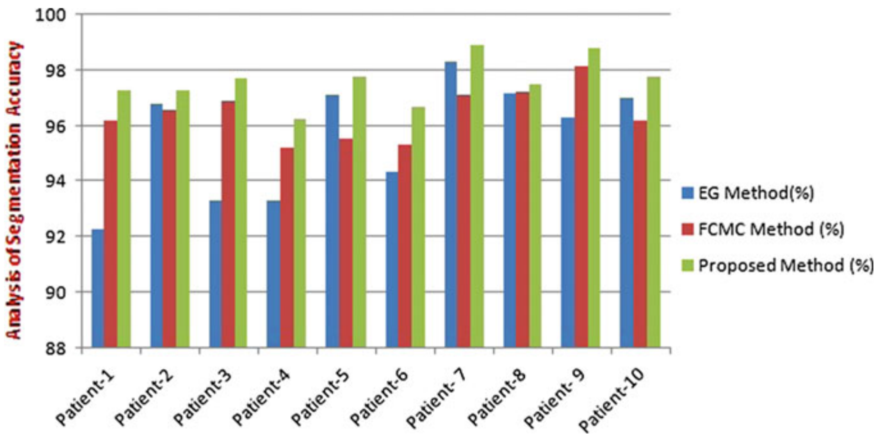


Fig. 4 Comparison of segmentation accuracy to proposed method, the improved Expectation Maximization (EM), and fuzzy C means clustering method

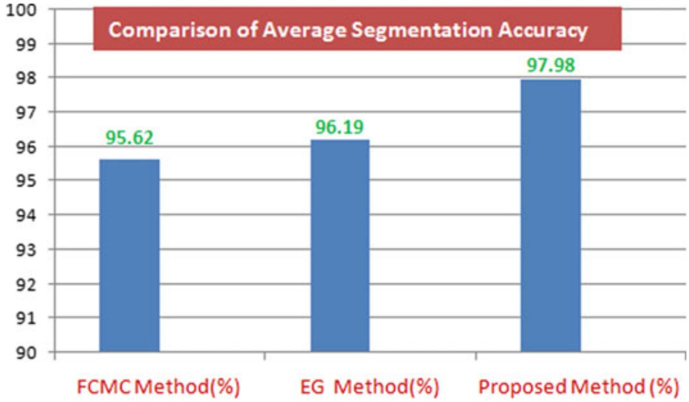


Fig. 5 Comparison of average segmentation accuracy

4 Conclusion

In this research paper, the wide-range of analysis has been done on numerous segmentation techniques and methods. The most popular contrast-enhanced T1C 3D brain magnetic resonance imaging modalities have been investigated for clinical tasks and detection of tumor. The proposed research work offered a contemporary fully automated segmentation framework for segmentation and detection of the brain tumor in contrast-enhanced T1C 3D brain magnetic resonance imaging which is successfully executed on BRATS simulated contrast-enhanced T1C 3D brain magnetic resonance imaging datasets and demonstrated the superior segmentation accuracy. The proposed segmentation method is evaluated on 10 patients' contrast-enhanced

T1C 3D brain magnetic resonance imaging dataset. The presented research work also demonstrated the relative analyses of various presented methods, i.e., fuzzy C means clustering method, improved Expectation Maximization (EM) method, and proposed method. The comparative study concludes that proposed method demonstrated improved segmentation accuracy than existing automated segmentation methods.

References

1. Bauer S, Wiest R, Nolte L-P, Reyes M (2013) A survey of MRI-based medical image analysis for brain tumor studies. Phys Med Biol 58(13):R97–R129
2. Weizman L (2012) Automatic segmentation, internal classification, and follow-up of optic pathway gliomas in MRI. Med Image Anal 16(1):177–188
3. Hameeteman K (2011) Evaluation framework for carotid bifurcation lumen segmentation and stenosis grading. Med Image Anal 15(4):477–488
4. Ahmed S, Iftekharuddin KM, Vossough A (2011) Efficacy of texture, shape, and intensity feature fusion for posterior-fossa tumor segmentation in MRI. IEEE Trans Inf Technol Biomed 15(2):206–213
5. Gooya A, Biros G, Davatzikos C (2011) Deformable registration of glioma images using EM algorithm and diffusion reaction modeling. IEEE Trans Med Imaging 30(2):375–390
6. Liu J, Li M, Wang J, Wu F, Liu T, Pan Y (2014) A survey of MRI-based brain tumor segmentation methods. Tsinghua Sci Technol 19(6):578–595
7. Palma CA, Cappabianco FAM, Ide JS, Miranda PAV (2014) Anisotropic diffusion filtering operation and limitations—magnetic resonance imaging evaluation. In: Preprints of the 19th world congress the international federation of automatic control, Cape Town, South Africa, 24–29 Aug 2014
8. Krissian K, Aja-Fernández S (2009) Noise-driven anisotropic diffusion filtering of MRI. IEEE Trans Image Process Arch 18(10):2265–2274
9. Bhima K, Jagan A (2017) An improved method for automatic segmentation and accurate detection of brain tumor in multimodal MRI. Int J Image Graph Signal Process (IJIGSP) 9(5):1–8. https://doi.org/10.5815/ijigsp.2017.05.01
10. Roerdink JBTM, Meijster A (2000) The fuzzy C means clustering method transform: definitions, algorithms and parallelization strategies. Fundam Inform 41:187–228
11. Li G (2010) Improved fuzzy C means clustering method segmentation with optimal scale based on ordered dither halftone and mutual information. In: 3rd IEEE international conference on computer science and information technology (ICCSIT), 9–11 July 2011, pp 296–300
12. Lin G-C, Wang W-J, Kang C-C, Wang C-M (2012) Multispectral MR images segmentation based on fuzzy knowledge and modified seeded region growing. Magn Reson Imaging 30(2):230–246
13. Gupta JMP, Shringirishi MM (2013) Implementation of brain tumor segmentation in brain MR images using k-means clustering and fuzzy c-means algorithm. Int J Comput Technol 5(1):54–59
14. Bhima K, Jagan A (2016) Novel technique for detection of anomalies in brain MR images. In: The international conference on frontiers of intelligent computing: theory and applications (FICTA-2016). Advances in intelligent systems and computing (AISC) Series. Springer, 16–17 Sept 2016
15. Stille M, Kleine M, Hagele J, Barkhausen J, Buzug TM (2016) Augmented likelihood image reconstruction. IEEE Trans Med Imaging 35(1):158–173

16. Nageswara Reddy P, Mohan Rao CPVNJ, Satyanarayana C (2016) Optimal segmentation framework for detection of brain anomalies. Int J Eng Manuf 6:26–37 (Published Online November 2016 in MECS)

Anatomical Segmentation of Human Brain MRI Using Morphological Masks

J. Mohamed Asharudeen and Hema P. Menon

Abstract Segmenting the anatomical parts of the human brain from MRI is a challenging task in medical image analysis. There is no evident scale for the distribution of intensity over a region in medical images. Region growing is performed to generate a mask to segment the anatomical parts from MRI. A new tailored version of dilation is used in renovating the segmentation mask. This custom-made dilation differs from typical dilation in computation. The structuring element size is fixed, and the anchor point norm is changed from usual dilation. The neighborhood evaluation is made only for certain pixels that are satisfied by the proposed constraints; thus, estimation is not made throughout the image. The computation of the classical dilation is reduced with the proposed custom-made dilation.

1 Introduction

Segmentation of a region in an image finds the homogenous similarity in the pixel intensities. The significant region of the image can be analyzed. Segmentation of brain anatomical parts is considered as one of the most critical steps in medical image analysis and it helps to examine the region with individual considerations. Since the regions of the brain image are difficult to observe from the gray level image segmentation. Equalization of the histogram over the image is made to make prominent regional differences. The image we have taken to examining the anatomical parts of the brain is the gray level image obtained from MRI Multiple Sclerosis Database (MRI MS DB) [1–4]. MRI determines the healthiness of the tissues present in the organs and growth of every tissue and also to measure the abnormal growth of tissues. MRI is used mostly in analyzing the tumors, brain injuries, and infec-

J. Mohamed Asharudeen (✉) · H. P. Menon
Department of Computer Science and Engineering, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Coimbatore, India
e-mail: cb.en.p2cvi16002@cb.students.amrita.edu

H. P. Menon
e-mail: p_hema@cb.amrita.edu

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_74

tions in the brain. MRI provides a detailed image of the brain. Brain lesions can be easily identified through MRI, as the appearance of the lesions will not be the same as a normal tissue. The interpretations of the brain parts are observed with color distributions in image processing. The hectic situation is there is no such scale for quantifying lesion intensity. The intensity is interpreted only through direct visual comparison with surrounding tissues.

2 Related Work

Hasan et al. [5] modified the existing morphological operations for 2D hole filling using dilation operation. The processing time of the algorithm reduced to one-third. Border image initial algorithm is modified by considering more starting points. This is considering the faster version of the existing hole-filling algorithm. Despotovic et al. [6] discussed the most popular methods used for brain parts segmentation in MR image, and comparison about them is also highlighted.

Guan et al. [7] proposed a novel model by integrating the Markov theory, the clustering theory, and normal curve model. Then the brain tissue has been extracted by the threshold segmentation. The work is compared with Markov algorithm and also, the algorithm of a 2D histogram with fuzzy clustering. Yazdani et al. [8] proposed a new technique for region dividing by combining the merits of region-based segmentation and the histogram-based image segmentation. This paper focused on the segmentation of the gray matter, white matter, and cerebrospinal fluid in brain MR images.

Somasundaram and Kalaiselvi [9] proposed an approach for filling the holes automatically in the binary masks based on run length encoded data. They claimed that the work is the best suitable holes-filling module for brain image segmentation. Piekar et al. [10] introduced the contrast agent for making the tumor visible. The region growing segmentation is employed. The region growing sometimes detects the healthy region as tumor boundaries. These overflow conditions are removed by using eight-neighborhood connectivity and specifies the importance of choosing the better homogeneity criteria for choosing the seed point.

Javadpour and Mohammadi [11] proposed a novel algorithm that uses regional growth methods and the seed points selections were automated by genetic algorithm for MRI segmentation. The proposed method claims that segmentation error can be reduced. Here, the standard deviation criterion is utilized to check the similarity. Mohd Saadd et al. [12] described the brain lesion segmentation. Lesion region is obtained by region splitting and merging. Automatic seed selection is performed using histogram thresholding that determines the optimal threshold value. Deng et al. [13] demonstrated the way to overcome the noises and threshold selection issues in tumor segmentation. An anisotropic diffusion filter is used to preserve the edges within which the proposed method chooses the mean–variance.

Roman-Alonso et al. [14] developed a tool called Data List Management Library used for segmentation of brain image volumes. They compared the proposed tool with other classical parallel approaches that are made with master–slave and static data distribution. Kapur et al. [15] presented a hybrid method for brain tissue segmentation from magnetic resonance images. They combined three existing techniques from the computer vision literature: expectation/maximization segmentation, binary mathematical morphology, and active contour models. They implemented the method on IBM's supercomputer Power Visualisation System and its results are validated by neuroanatomical experts. Park [16] presented the concept of mathematical morphological operations utilized for handling noises in tracking the objects.

3 Proposed Work

Region growing is one of the popular approaches of image segmentation [17]. It starts from the user interest to which region need to be segmented in an image. It plays a vital role in medical image analysis. The image we have taken to examine the anatomical parts of the brain is the gray level image obtained from MRI Multiple Sclerosis Database (MRI MS DB) [1–4]. The algorithm starts budding from a seed pixel. This method checks for the homogeneity criteria around the seed. Once the neighborhood connectivity is checked, the next search for the seed pixel is started around the previous results to avoid the region's growth along the dominant direction. The pixels traced by the above-described process will determine the region wants to get segmented. The flowchart for the segmentation technique is shown in Fig. 1.

There are many challenges faced during this segmentation work such as noises [18, 19]. Noises interrupt the growth along with the chosen region of importance. There are many chances of discontinuity of the segmented parts. Analysis of the partially segmented region may result in wrong diagnostics. Distributions of the colors are made uniform as a result of histogram equalization. The precise tissue layer is also getting segmented through this method. Every subtle layer of the cerebellum folds can be segmented. The segmentation process is made by region growing methodology with better precision and the binary mask is developed as its result.

3.1 Morphological Operations

Mathematical morphology in image processing is popular in handling noises, extracting the features in the images. These functions are working under the concept of set

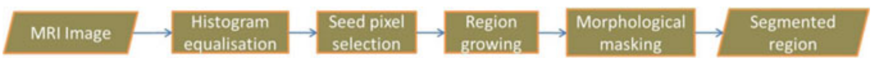


Fig. 1 Flowchart for brain parts segmentation

theory. The fundamental mathematical morphological operations are dilation and erosion.

These two primitive operations play a vital role in generating other morphological algorithms. The structuring element is a significant part in defining the shape of the object that required to be retained from the given image. The structuring element can be changed to any shape depends on the application.

Here in this proposed work, the structuring element is designed in an arrangement that it can be used for making the masks for the region to be segmented from human brain MRI. The typical dilation is applied in an inexplicable fashion for the fragmentary mask traced from the region growing technique. The dilation is tailored to two naïve operations, namely: waxing and waning. These operations are personalized to work with binary images.

Waxing: A structuring element B of size 3×3 has the anchor point in its origin and the waxing operation on the binary image A is performed only if the anchor point of the kernel has the intensity value 1. Here, the representation of waxing operation is chosen as $\oplus^{+}$. The anchor point of the kernel is checked and the individual neighborhood is changed only if the condition matched for waxing. The anchor point of the kernel is holding the intensity value as 1 and the conditions of the neighborhood hold the intensity value as 0 then replace them as 1.

$$A \oplus^{+} B = \left\{ x \in E | (\hat{B})_x \cap A \neq \emptyset, x_B = 1 \right\} \quad (1)$$

Let E be the Euclidean space and waxing is applied for $x_B = 1$, x_B be the anchor point of the kernel.

Waning: A structuring element B of size 3×3 has the anchor point in its origin and the waning operation on the binary image A is performed only if the anchor point of the kernel has the intensity value 0. Here, the representation of waning operation is chosen as $\oplus^{-}$. The anchor point of the kernel is checked and the individual neighborhood is changed if the condition is matched for waning. The anchor point of the kernel is holding the intensity value as 0 and the conditions of the neighborhood hold the intensity value as 1 then replace them as 0.

$$A \oplus^{-} B = \left\{ x \in E | (\hat{B})_x \cap A \neq \emptyset, x_B = 0 \right\} \quad (2)$$

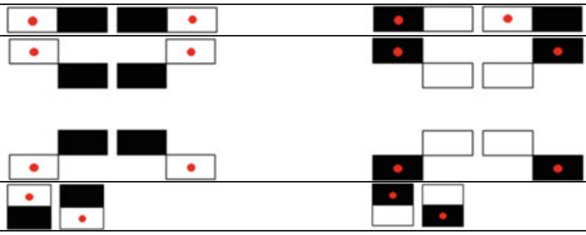
Let E be the Euclidean space and waxing is applied for $x_B = 0$, x_B be the anchor point of the kernel. The design of the structuring element is shown in Table 1.

These structuring elements are combined to form a 3×3 kernel. The red mark in Table 1 shows the anchor point of the individual structuring element. If the neighboring condition is satisfied the origin is changed its value. The combination of these structuring elements is shown in Fig. 2.

Figure 2 shows the designed kernel for waxing and waning as well as the existing kernel structures. The difference can be clearly visible from the red marks. The typical kernel structures will check for the neighborhood conditions and the anchor point located at the origin gets modified. The proposed kernel for waxing and waning

Table 1 (A) Horizontal structuring elements, (B) diagonal structuring elements, and (C) vertical structuring elements of waning and waxing kernels

Structuring element	Waning	Waxing
(A)		
(B)		
(C)		



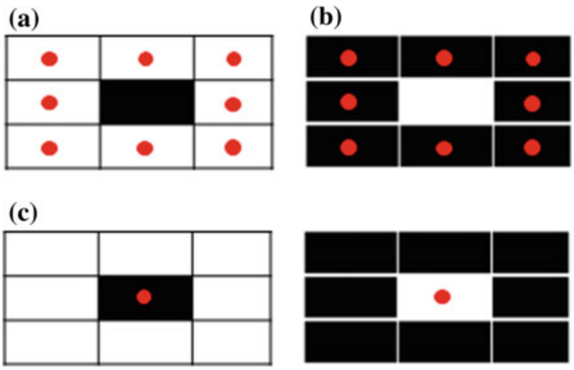


Fig. 2 **a** Waning kernel, **b** waxing kernel, and **c** existing 3 × 3 morphological kernel

will check the anchor point only then the neighborhood condition is checked. The modification is not made for the anchor point of the kernel. Instead, those pixels satisfy the neighborhood conditions. These two conventional dilation operations are named as waxing and waning.

The result of the region growing technique is a binary image. This binary image creates a checkerboard pattern as a result of four-neighborhood connectivity comparison. The four-neighborhood connectivity pattern for this mask construction used is “cross” arrangement as shown in Fig. 3. This mask undergoes waxing operation to fulfill the mask shape and it has been used to segment the region from MRI.

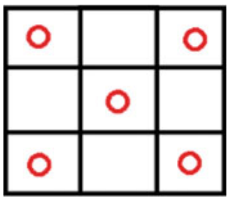


Fig. 3 Four-neighborhood connectivity “cross” design

3.2 Methodology

The proposed method uses a region growing technique to identify the structural parts of the brain. Once the histogram equalization is performed, the region of interest mask is prepared from the region growing technique. The mask is outlined with holes. The holes appeared due to four-neighborhood connectivity comparison. The results of segmentation are also made using eight-neighborhood connectivity as claimed by [8]. The eight-connectivity dominates the region's growth beyond the region of interest as shown in Fig. 5. The shape of the mask is treated with waxing operation. The final mask has been utilized to segment the region.

Let the structuring element of waxing is superimposed over the binary image. Check for the waxing constraints. Replacement of the pixel value is made in its neighborhood. Loop throughout the image. Waxing is used for objects shape expansion and waning is used for objects shape shrinking. The algorithm appends one-pixel width around the boundary of the object. The proposed work utilizes the four connectivity estimation. Waxing, a new method proposed to fulfill the shape of the region growing resultant mask in this proposal.

Let $I(x, y)$ be the given MRI image, and the corresponding mask for the ROI is $M(x, y)$.

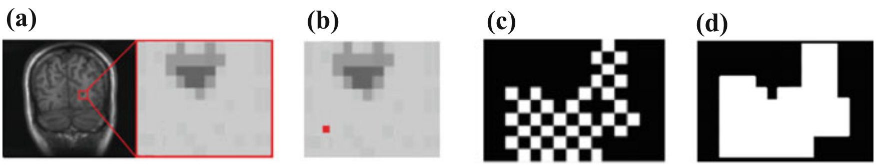


Fig. 4 **a** Region zoomed is examined with region growing, **b** red mark represents the seed pixel, **c** result of region growing technique, and **d** waxing output

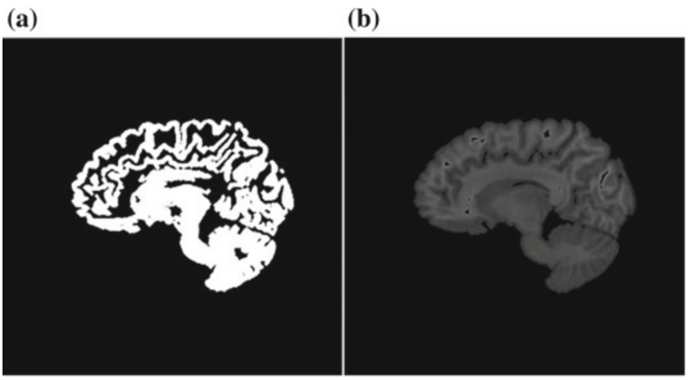


Fig. 5 Result of eight-neighborhood connectivity: **a** mask output and **b** segmented brain

$$M(x, y) = \begin{cases} 1, & N_4\{\nabla I(x, y)\} < T \\ 0, & \text{Otherwise} \end{cases}$$

Mask $M(x, y)$ undergoes *waxing* operation for filling the holes. With this mask, the region is segmented from the image $I(x, y)$.

$$M(x, y) \cap I(x, y) = R(x, y)$$

Let $R(x, y)$ be the segmented region.

Brain parts segmentation involves a region growing technique along with waxing operation as follows:

Step 1: Perform histogram equalization on the given image.
Step 2: Set the gradient value within the range of 0.00 and 0.01.
Step 3: Choose the region from the image, to get segmented by selecting any point in the region needs to be analyzed.
Step 4: Prepare the mask by selecting a seed pixel.
Step 5: Perform four-neighborhood connectivity evaluation from the selected seed pixel that is present in the mask.
Step 6: Mark those pixels that satisfy the constraints as white pixels in the mask.
Step 7: Loop through step 5 until the condition fails.
Step 8: Mask undergoes waxing and the result is shown in Fig. 4d.
Step 9: Segmentation of the region in the image is retrieved by the obtained mask.

4 Results and Analysis

The region growing is employed in brain segmentation for extracting the coherent intensity section and a preprocessing technique is also used for better localization of the region of interest. The eight-neighborhood connectivity is used, the region growing is dominated instead of the specified region of interest. The corpus callosum is selected as the region for segmenting from the given image. The result shows the entire brain from the raw MRI image as shown in Fig. 5. Eventually, the four-neighborhood connectivity is used for segmenting the definite region in the image. These results develop a mask for segmenting the region of interest from the original image. The created mask has some imperfect shape that is altered using waxing operation and the final mask is prepared. This mask will get superimposed over the original image for segmenting the definite region from the given image. The region growing process for corpus callosum from the sagittal slice is shown a successive growth step process in Fig. 6. The region growing algorithm using four-neighborhood connectivity is tested only using brain here. It can also be tested with some other organs for region segmentation.

The major difference between the standard dilation and these waxing and waning operations is: the origin differs as shown in Fig. 2, and the kernel matching is performed for replacing the anchor point of the kernel that satisfies the criteria. The neighborhood comparison is made throughout the image in the existing operation, but the proposed operation will perform the neighboring comparison only for the pixel that matches the anchor point of the kernel. The dimension is fixed as 3×3.

The result of the waxing and typical dilation will produce the same output as shown in Fig. 7 for the chosen structuring element. The major difference with these operations is the neighborhood pixel comparison shown in Table 2. The comparison of time complexity is shown in Table 3. This supports the claim that the proposed function works faster than the existing dilation operation. The proposed function has been verified with three categories of brain MRI.

Table 4 shows the anatomical parts of the brain from MRI of different slices. The result shows the partially generated mask and its zoomed view. The parts have been segmented through this method [20, 21] are evaluated for finding the similarity between the manually segmented and the proposed semiautomated segmented region of brain parts.

The values shown in Table 5 are almost nearer to 1 which represents the segmentation through this approach nearly matches the manual segments. The similarity

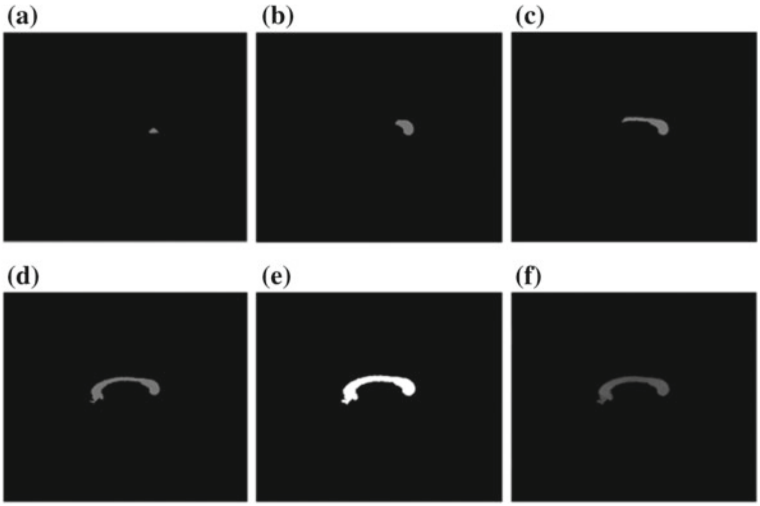


Fig. 6 **a**–**d** Successive region growths of corpus callosum, **e** final mask, and **f** segmented corpus callosum

Table 2 Neighborhood pixel comparison of waxing and dilation operation

Eight-neighborhood comparison	Waxing	Dilation
Total no. of pixels	80 pixels	192 pixels

measures [22–25] used here are SSIM and QIM. The structural similarity index is a method to compare the structure, contrast between the two images is calculated as in (3). Quality index measure compares contrast distortion, and luminance distortion between the images is calculated as in (4).

$$\mathrm{SSIM} = \frac{(2\mu_x\mu_y + c1)(2\sigma_{xy} + c2)}{(\mu_x^2 + \mu_y^2 + c1)(\sigma_x^2 + \sigma_y^2 + c2)} \tag{3}$$

$$\mathrm{QIM} = \frac{4\sigma_{xy}\overline{xy}}{(\sigma_x^2 + \sigma_y^2)(\bar{x}^2 + \bar{y}^2)} \tag{4}$$

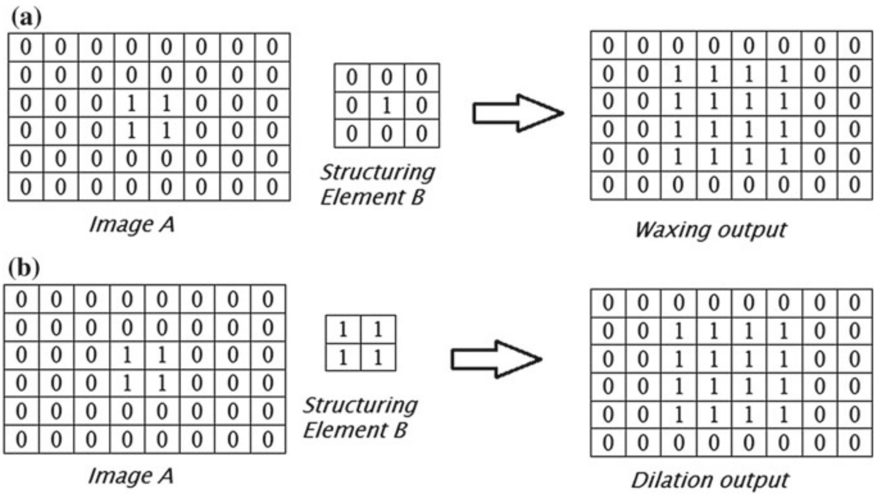
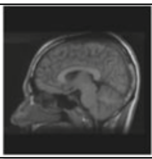
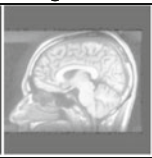
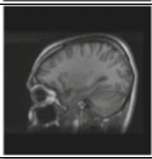
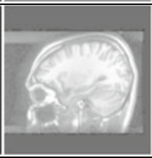
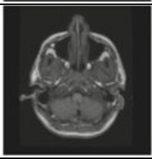
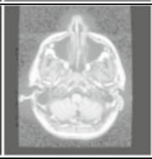
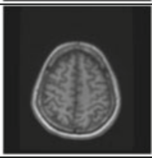
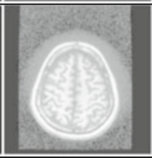
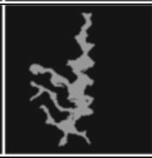
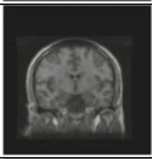
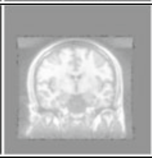
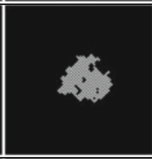
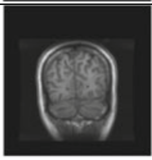
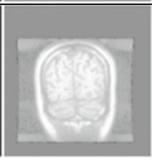


Fig. 7 **a** Waxing operation and **b** dilation operation

Table 3 Time taken for waxing and dilation operation

Masks developed from region growing	Time taken for waxing operation (s)	Time taken for dilation operation (s)
Sagittal_Corpus_Callosum	0.065	0.926
Sagittal_Arbor_vitae_of_cerebellum	0.073	0.997
Axial_Cerebellar_Tonsil	0.081	1.168
Axial_White_Matter	0.160	2.169
Coronal_Hippocampus	0.081	0.957
Coronal_White_Matter	0.074	1.003

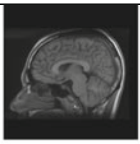
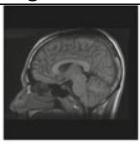
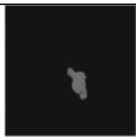
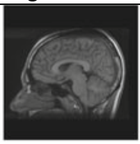
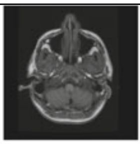
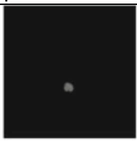
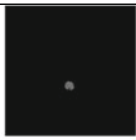
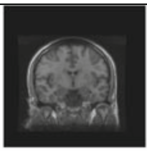
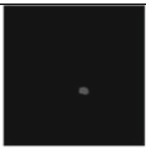
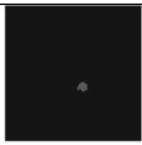
Table 4 Brain MRI segmentation results

Image slice	MRI image	Histogram equalised image	Region segmented from the brain	Zoomed view of the region segmented	Name of the region
Sagittal slice					Corpus callosum
Sagittal slice					Arbor vitae of cerebellum
Axial slice					Cerebellar tonsil
Axial slice					White matter
Coronal Slice					Hippocampus
Coronal Slice					White matter

5 Conclusion

The anatomical parts of the brain are segmented using the region growing methodology. The tailored mask is computed efficiently. The distinct way of dilation is employed in constructing the mask. This operation is also clarified with the existing dilation operation. The similarity measures are also shown the segmented results are almost similar to the manually segmented results. The segmented results show the better path for diagnosis of the anatomical parts of the brain. The region growing has

Table 5 Similarity measures between manual segmentation and proposed segmentation

Image Slice	Manual segmentation	Proposed segmentation	SSIM	QIM
 Sagittal slice	 Corpus collusum		0.9597	0.9418
 Sagittal slice	 Pons		0.9554	0.9373
 Sagittal slice	 Septum pellucidum		0.9612	0.9482
 Axial slice	 Spinal cord		0.9663	0.9441
 Coronal slice	 Thalamus		0.9667	0.9492

its own limitations in tracing the region of interest. If the histogram equalization hindered the region boundary identification, then the segmentation gives unpredicted results not appropriate for diagnostics. Preprocessing has to develop the vibrant boundary localization.

References

1. Loizou CP, Murray V, Pattichis MS, Seimenis I, Pantziaris M, Pattichis CS (2011) Multi-scale amplitude modulation-frequency modulation (AM-FM) texture analysis of multiple sclerosis in brain MRI images. IEEE Trans Inf Technol Biomed 15(1):119–129

2. Loizou CP, Kyriacou EC, Seimenis I, Pantziaris M, Petroudi S, Karaolis M, Pattichis CS (2013) Brain white matter lesion classification in multiple sclerosis subjects for the prognosis of future disability. Intell Decis Technol J (IDT) 7:3–10
3. Loizou CP, Pantziaris M, Pattichis CS, Seimenis I (2013) Brain MRI Image normalization in texture analysis of multiple sclerosis. J Biomed Graph Comput 3(1):20–34
4. Loizou CP, Petroudi S, Seimenis I, Pantziaris M, Pattichis CS (2015) Quantitative texture analysis of brain white matter lesions derived from T2-weighted MR images in MS patients with clinically isolated syndrome. J Neuroradiol 42(2):99–114
5. Hasan MM, Mishra PK (2012) Improving morphology operation for 2D hole filling algorithm. Int J Image Process (IJIP) 6(1):1–12
6. Despotovic I, Goossens B, Philips W (2015) MRI segmentation of the human brain: challenges, methods and applications. Comput Math Methods Med 1–23. Article ID 450341
7. Guan Y-H, Lv L, Duan R, Ji Y-H (2009) The brain segmentation by Markov field and normal distribution curve. In: Image signal processing, pp 1–5
8. Yazdani S, Yusof R, Karimian A, Mitsukira Y, Hematian A (2016) Automatic region-based brain classification of MRI-T1 Data. PLoS ONE 11(4):e0151326
9. Somasundaram K, Kalaiselvi T (2010) A method for filling holes in objects of medical images using region labeling and run length encoding schemes. In: National conference on image processing (NCIMP), pp 110–115
10. Piekar E, Szwarc P, Sobotnicki A, Momot M (2013) Application of region growing method to brain tumor segmentation-preliminary results. J Med Inform Technol 22:153–160
11. Javadpour A, Mohammadi A (2016) Improving brain magnetic resonance image (MRI) segmentation via a novel algorithm based on genetic and region growth. J Biomed Phys Eng 6(2):95–108
12. Mohd Saadd N, Abu-Bakar SAR, Muda S, Mokji M, Abdullah AR (2012) Automated region growing for segmentation of brain lesion in diffusion-weighted MRI. In: Proceedings of the international multi conference of engineers and computer scientists, vol. I, pp. 14–16, 674–677, March 2012
13. Deng W, Xiao W, Deng H, Liu J (2010) MRI tumor segmentation with region growing method based on the gradients and variances along and inside of the boundary curve. In: 3rd international conference on biomedical engineering and informatics, vol 1, pp 393–396
14. Roman-Alonso G, Jimenez-Alaniz JR, Buenabad-Chavez J, Castro-Garcia MA, Vargas-Rodriguez AH (2007) Segmentation of brain image volumes using the data list management library. In: 29th annual international conference of the IEEE engineering in medicine and biology society, pp 2085–2088
15. Kapur T, Grimson WEL, Wells WM III, Kikinis R (1996) Segmentation of brain tissue from magnetic resonance images. Med Image Anal 1(2):109–127
16. Park H. A method of controlling mouse movement using a real time camera. Department of Computer Science, Brown University
17. Pohle R, Toennies KD (2001) A new approach for model-based adaptive region growing in medical image analysis. In: International conference on computer analysis of images and patterns, pp 238–246
18. Hiralal R, Menon HP (2016) A survey of brain MRI segmentation methods and issues involved. In: International symposium on intelligent systems technologies and applications (ISTA'16), Advances in intelligent systems and computing
19. Nilakant R, Menon HP, Vikram K (2017) A survey on advanced segmentation techniques for brain MRI image segmentation. Int J Adv Sci Eng Inf Technol 7(4)
20. Bhuvana D, Bhagavathi Sivakumar P (2014) Brain tumor detection and classification in MRI images using probabilistic neural networks. In: Proceedings of the second international conference on emerging research in computing, information, communication and applications (ERCICA-14). Elsevier, pp 796–801
21. Renjini H, Bhagavathi Sivakumar P (2013) Comparison of automatic and interactive image segmentation methods. Int J Eng Res Technol (IJERT) 2(6):3162–3170

22. Wang Z, Bovik AC (2005) Structural approaches to image quality assessment. In: The handbook of image and video processing, 2nd edn. Academic Press, New York, pp 961–974
23. Wang Z, Bovik AC, Sheikh HR, Simoncelli EP (2004) Image quality assessment: from error visibility to structural similarity. IEEE Trans Image Process 13(4):600–612
24. Wang Z, Bovik AC (2003) Objective video quality assessment. In: The handbook of video databases: design and applications. CRC Press, Boca Raton, pp 1041–1078
25. Wang Z, Bovik AC, Lu L (2002) Why is image quality assessment so difficult? In: IEEE international conference on acoustics, speech, and signal processing

Digital Image Restoration Using NL Means with Robust Edge Preservation Technique

Bhagyashree V. Lad, Bhumika A. Neole and K. M. Bhurchandi

Abstract We present a novel approach for image denoising with spatial domain edge preservation method to yield denoised image, from an image which is distorted by additive white Gaussian noise without loss of detail information of an image. Denoising of noise corrupted images while preserving its attributes like edges and fine details is an extreme challenge. During image acquisition and transmission often noise gets added in the image which degrades the image quality. Algorithms like NL means and BM3D have been very successful in this aspect. The weight assigned to the fellow pixels for calculating replacement for a noisy pixels have a pivotal role in these algorithms. In NL means, the weight assigned to a pixel depends only on its relative magnitude with respect to its neighborhood. In this paper we are proposing a robust mechanism for weight calculation which also considers the orientation of a pixel. Using this mechanism we got improvement in PSNR and visual quality of the de-noised image as compared to the NL means method and also got improvement in structural similarity as compared to the BM3D. From the experimental results it can be seen that our proposed algorithm is superior in comparison to the NL means and BM3D technique of image denoising, considering the factors like PSNR, SSIM and accordingly conclusions are drawn.

B. V. Lad (✉) · B. A. Neole
Department of Electronics and Communication Engineering, Shri Ramdeobaba College of Engineering and Management, Nagpur 440013, India
e-mail: ladbv@rknec.edu

B. A. Neole
e-mail: neoleba@rknec.edu

K. M. Bhurchandi
Department of Electronics and Communication Engineering, VNIT, Nagpur 440010, India
e-mail: bhurchandikm@ece.vnit.ac.in

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_75

1 Introduction

Noise is an unsystematic and indeterminate variation in the intensity of the image. All the images contain certain amount of noise. An image captures a moment forever. Today there is almost no area of technical endeavor that is not impacted in some way by digital images and hence their processing. There are many fields where decisions are made solely based on the digital images as inputs. Then it becomes very important that these images are noise-free and pure. Thus this is the challenge that we need to obtain best possible approximation of original image from available degraded image. Many a times the noise gets added in an image while capturing it by the camera. Images are also contaminated by the noise while capturing the images through Charged Coupled devices (CCD), CMOS Sensors and Contact Image Sensors (CIS). Some signals required for TV to broadcast the data which is transmitted over cable or received at the antenna and in case of digital images the signal is the light which hits the camera sensor. In this transmission process the noise gets added in the signals. This noise appears as random signals which degrade the image quality. Thus, it is a great challenge for the researchers today to remove noise from images. There are various image denoising algorithms approaches which have been employed such as Perona and Malik proposed a scale space-based approach [1], Nobert Wiener presented an adaptive filter [2], furthermore several versions of wavelets [3–5] and spatial filters [6–9] are used and yielded interesting result to remove the noise. Every algorithm has certain merits and demerits.

There is a general belief that noise is a high-frequency component and the signal is a low-frequency component. An image itself has high-frequency attributes like edges which are sharp change in intensity and fine details which may get misinterpreted as high-frequency noise component by the de-noising algorithms and thus ruined.

An image can be considered as noisy if its pixel values do not match with the pixel values in some reference original image. In a way all the digital images are noisy as the process of quantization itself introduces the quantization noise. Further, noise may be introduced during any stage like while acquiring an image or maybe while transmitting it. The nature of this noise can be modeled by various noise models like salt and pepper noise, speckle noise, Gaussian noise, periodic noise etc. Out of these, generally, the independently and individually distributed additive white Gaussian noise can be considered as the better approximation to the practical noise.

Many of the methods used for denoising work poorly in case of recovery of edges. The edges become blurred because of averaging. Mostly, the noise present in digital images is additive in nature in which the uniform power is distributed in whole bandwidth and the noise generally has the Gaussian probability distribution. Such a noise is generally called as Additive White Gaussian Noise (AWGN). Because of poor image acquisition techniques and poor transmission through noisy communication channel the White Gaussian noise gets introduced in an image. In most of the image denoising algorithms the artificially distorted images by White Gaussian noise are considered to get the test results [3–7]. It can be seen that Gaussian filter even though has desirable features, the satisfactory results are not obtained because of edge detection techniques involved in the denoising algorithm [8].

Images are also affected by salt and pepper noise for which median filter is to be used for removal of such noise. Median Filter is an influential nonlinear filter which is one of best and simple method in spatial domain. But it is efficient only for low noise densities and yield blurred images at high noise densities though it is an easy method to perform image smoothing. Median filter reduces the intensity variation between the neighboring pixels. In median filtering technique we replace the pixel value of an image with the median of all the neighboring pixel values instead of replacing it with the mean of neighboring pixels. To calculate the median value, all the pixel values are sorted in the ascending order first and the pixel is replaced by the middle pixel value of the sorted sequence. If the sorted sequence consists of even number of pixels then the desired pixel is replaced by the average of two middle pixel values in the sequence. The median filter yields best result when percentage of impulse noise present in an image is less than 0.1%. When the impulse noise is present in large amount then the median filter does not give best result because it tends to remove image details at the time of removing noise such as lines and the corners and thus degrades the performance in case of signal-dependent noise.

The most popular and well-known Wiener filter can also be used for removal of image degradations. The Weiner filter, which is based on the statistical approach, filters the noise which has degraded the signal. The designing of all the filters is mostly based on desired frequency response, but the designing of Weiner filter takes the different approach. Weiner filter is designed to reduce the mean square error to the most extent. So, this filter more effectively reduces the noise and various degrading functions.

De-noising techniques can be implemented in spatial domain or in frequency domain. The non-local means and the BM3D are two very successful de-noising methods. The non-local means is a pure spatial method while the BM3D is implemented partly in spatial and partly in frequency domain.

An isotropic Gaussian kernel is used for denoising in NL means algorithm. Kernels provide NL means with more similar patches, in which maximum similar patches are obtained near the edges where the contrast is very high. Depending upon the patch size two drawbacks arise. By choosing the small patch size noise remains in homogeneous parts of image, on the other side, choosing the large patch size leads the noise to halo near the edges. All these situations happen when the patches that are similar to the current patch are rare. Some attempts have been made to remove noise halo, but a better technique can be used which is based on shaped and directional patches. Number of iterations are required to implement this technique so computational time increases and finally SURE criterion is used to choose the best kernel for each part of image. Although this method improves NL means performance by eliminating the noise halo, but it gives rise to some drawback. The first drawback is selection of limited number of predefined kernels which leads to difficulty in denoising the images with different structures. Also an image consists of a large number of different structures and textures which requires a huge number of predefined kernels. The second drawback is the algorithm becomes computationally inefficient because of big group of kernels which leads to numerous iterations. And the third drawback is due to lack of robustness of SURE. These drawbacks can be eliminated by using Adaptive NL means method.

The BM3D algorithm presents the redundant successive computation of transform of similar overlapping image patches and their so-called 3D transform computations followed by collaborative filtering and inverse 3D transform. Averaging of the overlapping redundant inverse transformed image patches is used to have a final pixel estimate. This is followed by a Wiener filter.

In this paper we are modifying subtly the weight calculation mechanism used in the non-local means algorithm. For finding a replacement for a noisy pixel, in non-local means, the idea is to assign weights to other pixels depending on the similarity of their neighborhood in terms of magnitude of the pixels with the neighborhood of the pixel which is under consideration. We propose dependence of the weight assigned to a pixel by considering adaptive kernel which is not only based on the magnitude of the pixels in its neighborhood but also on the orientation of that pixel.

The paper is divided in 5 sections. Section 1 briefs the noise model considered. Section 2 explains the non-local means algorithm, whereas the concept of determining orientation of a pixel is given in Sect. 3. Section 4 demonstrates the modified algorithm and finally in Sect. 5, the results are shown and finally the conclusion is drawn.

2 Methods and Materials

2.1 Noise Model

Let X be the original image of size $m \times n$ and let N represent the additive white Gaussian noise matrix of the same size. Then the corrupted image can be given by Y as

$$Y = X + N \tag{1}$$

Since the noise model is additive, the value at every point in the corrupted image is simply the addition of the corresponding values in the original image and the noise matrix. The values in this nose model have a Gaussian distribution [14].

2.2 Non-local Means

Antoni Buades presented the algorithm for noise removal that is non-local means filter which has received a lot of attention from image processing community [9, 10]. Non-local filtering considers all pixels for finding replacement of one pixel. Weights are assigned to all the pixels depending on their similarity with pixel to be replaced that is NL means uses the self-similarity concept. In this algorithm each pixel is represented by the weighted average of all the pixels in the neighborhood of the image. To compute similarity between two neighborhoods the weighted sum of the squares of difference between the two neighborhoods is calculated. The non-local means algorithm considers a small patch around a pixel. It doesn't consider only the central patch. The non-local means filter gives more weight to the pixels which are there in the center of neighborhood and less weight to those pixels which are near to the edges.

Thus taking the base of this NL means algorithm, in this paper we have further improved the image quality and compared the results with different filtering techniques and tried to bring the results close to the results obtained by the state of the art techniques. The image qualities are compared on the basis PSNR (Peak Signal to Noise Ratio) and SSIM (Structural Similarity Index Measure).

In NL-means algorithm the pixel intensity is represented as a sum of weighted Gaussians of each neighboring pixel. This algorithm is computationally very heavy due to the involved exponential computations in the neighborhood of all pixels. Also calculation of the weights is computationally heavy. It also blurs the images considerably if exact weights are not selected [11, 12].

To determine the weights of the filters for filtering the noisy images, non-local means filter uses all the possible self-predictions and self-similarities which the image imparts provided the image contains large amount of self-similarity. Because of the self-similarity in an image the pixels are highly correlated and because of the independent and identical distribution property of the noise, the averaging of such highly correlated pixels suppresses the noise which yields to the pixel which is nearly identical to its original value. For a given noisy image $v = \{v(i)|i \in I\}$, where I is a set of the whole pixels in the image, and i is an arbitrary pixel in I. The pixel i is processed by the non-local means filtering and the formula can be given by

$$\mathrm{NL}[v](i) = \sum_{j \in I} w(i, j)v(j) \tag{2}$$

Where $w(i, j)$ is the weight between i and j, and it satisfies $0 \le w(i, j) \le 1$,

$$\sum_{j \in I} w(i, j) = 1 \tag{3}$$

By using this formula we can get the intensity of pixel i after denoising. Each pixel is expressed as the weighted average of all the pixels in an image.

2.3 BM3D

Another method which is most popularly used for image denoising is BM3D. In BM3D 3D arrays are formed by collecting the similar 2D image blocks. These 3D data arrays are called as groups. The filtering process used for these 3D groups is collaborative filtering. Collaborative filter gives the finest details of the grouped blocks and it also preserves all the features of an individual block. After this process of filtering, the blocks are returned to their original position [13].

BM3D, i.e. Block matching 3D generally deals with the improved thinly scattered data representation in the transform domain. And this improvement in the sparsity can be achieved by the formation of 3D data arrays by grouping the nearly identical 2D fragments in an image. For processing these 3D data arrays a special technique of the Collaborative filtering is used. This technique consist of the steps like, performing 3D transformation on 3D data array called as groups, then carrying out the shrinkage of the transformed coefficients and finally retrieving the image by applying inverse 3D transform to the obtained coefficients. Thus, a 3D estimate of the group is obtained which is nothing but the combination of filtered 2D image blocks. As the grouped fragments are highly similar, this technique gives the sparse representation of the true signal. And so the noise can be disarticulated by using the shrinkage technique. So the collaborative filtering divulges the sharp and fine details in conjunct with the groups. And also it preserves all the unique features of each individual grouped fragment. The grouping is done here so that the higher dimensional filters can be used for each group, which then utilizes the similarity between each data groups so that the true signal can be estimated for each of the data array.

In BM3D, block matching is used to perform the grouping of data and then the collaborative filtering is used which is achieved by shrinkage of 3D transform domain coefficients. In general the image group fragments are fixed sized square blocks. The generalized BM3D algorithm can be represented as follows. The reference blocks are extracted from the input noisy image and the following process is applied to each sub-block.

- Collect the blocks in a stack which are similar to the reference block to form a 3D data array, i.e. groups. This observation of the similar blocks to the reference one is called as block matching.
- Carry out the collaborative filtering of the group and put the estimated 2D data of all the grouped blocks to their original locations.
- Because of the overlapping of the obtained blocks during processing, there can be multiple estimates for each pixel. So average of all these estimates belong to one pixel and replace the original pixel value by this aggregate. In this way, the whole image can be estimated.

3 Proposed Algorithm

In this work, we have modified subtly the non-local means algorithm by using the concept of block matching from the BM3D technique. For finding a replacement of a noisy pixel in non-local means, the idea is to assign weights to other pixels depending on the similarity of their neighborhood with the neighborhood of the pixel which we are currently processing. Here, the similarity of the neighborhoods is computed in terms of their weighted averages. Instead of using weighted averages if the neighborhoods are matched as blocks, we obtain good results than classical non-local means algorithm.

The concept of block matching improves the performance of non-local means. In NL means the weight of a pixel (candidate pixel) depends only on how similar is its weighted average with respect to the weighted average of the pixel (central pixel) whose estimate is being calculated. More closely their weighted averages, more is the weight or dominance of the candidate pixel in calculating an estimate. If along with the weighted average, the similarity in the orientation of the candidate and the central pixel is also considered for calculation of weight, the estimated value of the central pixel gets even more close to its true value. In non-local means based on edge detection, the weights are calculated considering the similarity in both the weighted averages and the orientation of the candidate and the central pixel.

To enumerate the similarity between two neighborhoods we have to take the weighted sum of squares of difference between the two neighborhoods. The non-local means filter gives more weight to those pixels which are there in the center of the neighborhood and less weight to those pixels which are near the edges.

Here, the replacement value of a pixel completely depends on the other pixels in that image. The more similar a pixel is to the pixel for which replacement is to be found out, the more its weight is. Estimation of a corrupted pixel is affected by the pixels having higher weights.

Consider a noisy pixel p. For this noisy pixel, all the other pixels in the corrupted image will be considered for determining an estimate. Let the other pixels be denoted by qi, where $i = 1$ to $N - 1, N =$ total number of pixels in the given image. Let weights be denoted by $f(p, qi)$, i.e. weight of $q1 = f(p, q1)$, weight of $q2 = f(p, q2)$ and so on. Then, the estimate for pixel p denoted by $\hat{p}$ is given by the expression

$$\hat{p} = (q1 * f(p, q1) + q2 * f(p, q2) + q3 * f(p, q3) + \ldots)/c_p \tag{4}$$

$$c_p = \text{normalizing factor} = \sum f(p, qi) \tag{5}$$

3.1 Calculation of Weight

For calculation of $f(p, q1)$, i.e. weight for pixel $q1$ w.r.t p, we need to consider a neighborhood of $q1$ of size $n \times n$ where n is some odd number greater than or equal to 3, preferably 3. Let *hq* denote the average of neighborhood again of size $n \times n$ centered at $q1$ and *hp* denote the average of neighborhood of p, then

$$f(p, q1) = \mathrm{e}^{\frac{-(hp-hq)^2}{sf}} \tag{6}$$

where, *sf* corresponds to smoothing factor, e.g. $sf = (10 * \text{sigma})^2$ where, sigma represents the standard deviation of the noise model added. Once we have calculated weights for all pixels, we can calculate the normalizing factor c_p and hence an estimate for pixel p. But computationally this algorithm is very complex and heavy. We can make it computationally efficient by incorporating the following steps:

- Instead of considering all the remaining pixels to find out replacement for a particular pixel, define a search window symmetric about the pixel which we have considered. In this way, only those pixels which are there within the search window of the given pixel and not all pixels of the image will influence the value of its estimate.
- Consider only those pixels about whom the search window fits properly. This will obviously leave the boundary pixels unchanged which can be omitted from the image later. Also within a search area the candidate pixels would be those about whom the $n \times n$ window fits.

These steps are rather rational as one expects a pixel most likely to be similar to a small block of pixels around it which forms its search area.

3.2 Algorithm

- Consider the noisy image.
- Define the window size and the search area.
- Start from the pixel about which the search area fits properly.
- Within the search area of current pixel (say p), start from such a pixel (say q) about which the window fits properly.
- To calculate weight for q ($f(p, q)$), we need hp = weighted average of window of pixel p, hq = weighted average of window of pixel q, rx = orientation of window of p along x-direction, ry = orientation of window of p along y-direction, ed = overall orientation of window of p, $rx1$ = orientation of window of q along x-direction, $ry1$ = orientation of window of q along y-direction and $ed1$ = overall orientation of window of q.
- $f(p, q) = \mathrm{e}^{\left(-\frac{(hp-hq)^2}{(\text{smoothning factor}\times\text{sigma})^2} - \frac{(ed-ed1)^2}{(\text{smoothning factor}\times\text{sigma})^2}\right)}$
- Similarly, find the weights for the other pixels in the search area of p.
- Calculate a normalizing factor, $c_p = \sum_i f(p, q_i)$
- Estimation of pixel $p = \frac{\sum_i q_i \times f(p, q_i)}{cp}$
- Repeat this procedure of all such valid p pixels.

Edge detection has been carried out by Sobel gradient operators along X-direction and Y-direction [14]. After Edge detection Wiener filter is also applied to denoised image to minimize the mean square error between estimated random process and desired process. So the experimental results demonstrate that our algorithms performance is superior to the NL-means and comparable with BM3D in terms of PSNR and SSIM.

4 Experimental Results

Table 1 presents the de-noising performance of the above algorithm along with NL means, BM3D methods. Zero mean Gaussian noise with standard deviation values of 10, 20, 25, 30, 35, 50, 80 and 100 have been introduced in the original gray level images of Cameraman of size 256×256, Lena and Boat of size 512×512. Peak signal to noise ratio (PSNR) in dB is chosen as the performance measuring parameter. More the value of PSNR more the removal of noise is. More value of PSNR indicates better results of the algorithm. It can be clearly observed that the PSNR values of the images de-noised by the proposed algorithm are greater than the PSNR values of the same set of images when de-noised by NL means.

As the standard deviation of the noise increased, the performance of the existing methods lowered as expected but the performance of the proposed algorithm still had an edge over that of NL means. In our proposed algorithm we experimented with the size of search area from 7×7 till 21×21 for a particular standard deviation of noise. It was found that for a given standard deviation, the PSNR first increased as the search area was widened but then after the search area reached to an extent, PSNR kept on dropping gradually.

On an average the PSNR increased until the search area was increased to 11×11. Beyond this point the PSNR mostly decreased. So the optimum search area according to the experimentation should be 11×11 with window size of 3×3. Visually the difference between the outputs of the two algorithms is indiscernible but mathematically, our proposed algorithm performs better than the NL means. Below table shows the denoised comparison of classical non-local means filtering, BM3D and proposed method.

In Table 1, consider denoising of corrupted image of Cameraman with Gaussian noise having standard deviation 10 which has PSNR of 28.13 dB. NL means improved the PSNR by 4.54 dB, making it 32.67 dB. BM3D improved the PSNR by 5.95 dB, making it 34.08 dB. Our algorithm improved it by 5.41 dB, making it 33.54 dB. As the standard deviation was increased, the PSNR of the corrupted image kept on decreasing. But the amount by which the PSNR of denoised image was lifted kept on increasing. This implies that more the noise gets added to the original image, the more noise is also removed by these algorithms and their performance thus increases. When the standard deviation was 50, the corrupted image's PSNR was 14.15 dB. BM3D uplifted this PSNR by 11.69 dB(against only 5.95 dB when standard deviation was 10), NL means improved the PSNR by 10.19 dB making it 24.34 dB and proposed

Table 1 De-noising performance of the proposed algorithm along with the performance of NL means and BM3D on gray images

Images	Sigma/PSNR	Non-local means		BM3D		Proposed algorithm	
Cameraman (256 × 256)		PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
	10/28.13	32.67	0.90	34.08	0.93	33.54	0.92
	20/22.08	29.32	0.82	30.39	0.87	29.85	0.85
	25/20.15	27.35	0.77	29.45	0.85	28.96	0.82
	30/18.55	26.72	0.74	28.64	0.83	28.08	0.80
	35/17.55	26.09	0.71	27.93	0.82	27.30	0.79
	50/14.14	24.34	0.61	25.84	0.78	25.23	0.75
	80/10.04	22.2	0.48	24.05	0.71	23.73	0.69
	100/8.12	20.59	0.39	22.81	0.67	22.17	0.64
Lena (512 × 512)	10/28.13	33.28	0.85	35.93	0.96	35.13	0.95
	20/22.11	30.98	0.78	33.05	0.94	32.50	0.92
	25/20.16	29.84	0.80	32.08	0.92	31.64	0.90
	30/18.58	29.28	0.72	31.26	0.91	30.82	0.87
	35/17.22	28.49	0.69	30.47	0.89	29.98	0.84
	50/14.14	26.72	0.60	28.86	0.86	28.21	0.82
	80/10.06	24.32	0.47	27.02	0.80	26.53	0.78
	100/8.14	22.46	0.37	25.57	0.76	24.95	0.74
Boat (512 × 512)	10/28.13	32.25	0.84	33.92	0.96	33.17	0.95
	20/22.11	28.68	0.77	30.88	0.92	30.16	0.91
	25/20.16	27.44	0.71	29.91	0.90	29.30	0.89
	30/18.58	26.56	0.67	29.12	0.88	28.82	0.86
	35/17.22	25.80	0.61	28.43	0.86	27.90	0.84
	50/14.14	24.27	0.57	26.64	0.81	26.06	0.77
	80/10.06	22.62	0.45	24.74	0.72	24.29	0.69
	100/8.14	21.94	0.40	23.74	0.68	23.15	0.65

algorithm outperformed NL means by improving the PSNR by 11.08 dB making it to 25.23 dB.

In case of Lena, though the performance of BM3D has been the best but the PSNR obtained from the proposed algorithm is majorly more than that obtained from NL means. So, here our proposed algorithm clearly outperforms NL means and nearer to BM3D.

Figure 1 shows the comparative denoised visual results of Lena and Cameraman corrupted with Gaussian noise of standard deviation 25 by NL means, BM3D and proposed algorithm. Figures 2a, b represents the original zoomed image of Lena eye and the respective image added with zero mean Gaussian noise with standard

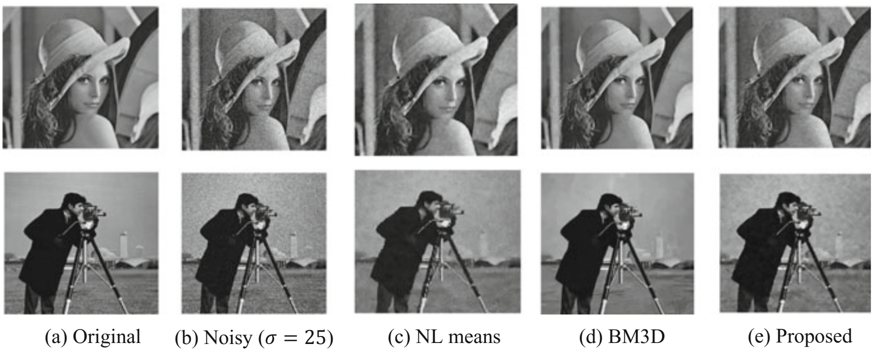


Fig. 1 **a** Original images of Lena and Cameraman, **b** Respective noisy images with $\sigma = 25$, **c**, **d**, **e** denoised images by NL means, BM3D, proposed algorithm respectively

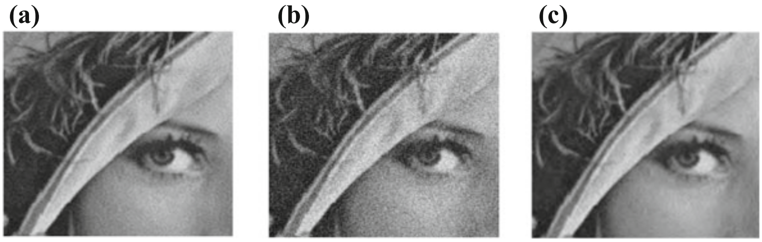


Fig. 2 **a** Original image of Lena eye . **b** Respective noisy image with $\sigma = 25$ and **c** denoised image of Lena eye by proposed algorithm

deviation 25 and Fig. 2c shows the quantitative results of the proposed algorithm. As it is visually difficult to observe the actual denoising for the whole image here, so we are presenting the results on zoomed portion of Lena eye. It can be seen that the PSNR of Lena eye is increased from 20.1 to 34.25 dB using the proposed algorithm with respect to noisy image.

5 Conclusion

A novel denoising technique for minimizing Gaussian noise using NL means and spatial domain edge preservation is presented in this work. The proposed algorithm gives good denoising performance as well as preservation of fine details and sharp edges of an image. The proposed algorithm yields better PSNR for low and moderate magnitudes of zero mean Gaussian noise. The recursive application of the algorithm is time-consuming like any other denoising algorithm. The algorithm is showing remarkable denoising capability of spatial domain edge preservation compared to other contemporary algorithms.

References

1. Perona P, Malik J (1990) Scale-space and edge detection using anisotropic diffusion. IEEE Trans Pattern Anal Mach Intell 12(7):629–639
2. Wiener N (1949) The extrapolation, interpolation and smoothing of stationary time series with engineering applications. Wiley, New York
3. Yaroslavsky LP (1985) Digital picture processing—an introduction. Springer, New York
4. Coifman RR, Donoho DL (1995) Translation-invariant de-noising. In: Antoniadis A, Oppenheim G (eds) Wavelets and statistics. Springer, New York, pp 125–150
5. Starck J, Candes EJ, Donoho L (2002) The curvelet transform for image denoising. IEEE Trans Image Proc 11(6):670–684
6. Smith SM, Brady JM (1997) SUSAN-A new approach to low level image processing. Int J Comput Vis 23(1):45–78
7. Buades A, Coll B, Morel JM (2005) A non-local algorithm for image denoising. In: Proceedings IEEE Computer society conference on computer vision and pattern recognition, San vol 2, pp. 60–65, Diego, USA
8. Liu Y-L, Wang J, Chen X (2008) A robust and fast non-local means algorithm for image denoising. J Comput Sci Technol 23(2):270–279
9. Buades A, Coll B, Morel J (2004) Technical report on image denoising methods: 2004–15, CMLA
10. Buades A, Coll B, Morel JM (2005) A review of image denoising algorithms, with a new one. Multiscale Model 4(2):490–530
11. Deledalle CA, Duval V, Salmon J (2011) Non-local methods with shape-adaptive patches (NLM-SAP). J Math Imag Vis 1–18
12. Tasdizen T (2009) Principal neighborhood dictionaries for nonlocal mean image denoising. IEEE Trans Image Proc 18:2649–2660
13. Dabov K, Foi A, Katkovnik V, Egiazarian K (2007) Image denoising by sparse 3-D transform-domain collaborative filtering. IEEE Trans image process 16(8):2080–2095
14. Gonzales RC, Woods RE (2012) Digital image processing, 3rd edn. Pearson Education Inc, pp 350–404

Speech Recognition Using Novel Diatonic Frequency Cepstral Coefficients and Hybrid Neuro Fuzzy Classifier

Himgauri Kondhalkar and Prachi Mukherji

Abstract Speech recognition is the ability of the machine to identify spoken words and classify them into appropriate category. First stage in the process of speech recognition is the extraction of appropriate features from the recorded words. We propose a novel algorithm for feature extraction using diatonic frequency cepstral coefficients. Diatonic frequencies are derived from a musical scale called as diatonic scale. The scale is based on harmonics of sound and models nonlinear behavior of human auditory filter. After feature extraction, the next classification stage uses a hybrid classifier using artificial neural network and fuzzy logic. If the difference between prediction values available at the output of the neural network is less, the classifier matches wrong patterns. Proposed algorithm overcomes this drawback using fuzzy logic. Proposed hybrid classifier improves the recognition rate significantly over existing classifiers. Test bed used in the experimentation focuses on Marathi language. It is the native language spoken in the state of Maharashtra.

1 Introduction

Spoken numeral recognition can be used for real-life applications like automated teller machines, lifts for visually sighted people. It can also be used for hands-free dialing systems for phones, search engines, etc. An efficient algorithm for both feature extraction stage and modeling stage of spoken word recognition is proposed in this work. Characteristics of sound are loudness, pitch, and quality. The speech signal can be represented using different types of parametric representations based on these characteristics. The existing techniques for the same are Mel frequency cepstral coefficients (MFCC) and Bark frequency cepstral coefficients (BFCC). Mel scale is a

H. Kondhalkar (✉)
Sinhgad College of Engineering, Pune, India
e-mail: gouri.ghule@viit.ac.in

P. Mukherji
Cummins College of Engineering, Pune, India
e-mail: prachi.mukherji@cumminscollege.in

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_76

perceptual scale based on "pitch" measurement of the signal. The cutoff frequencies used in Mel scale model the nonlinear perception of frequencies in the human auditory system [1]. The maximum sampling frequency suggested for Mel scale is 14,000 Hz. Human hearing range is divided into 24 nonoverlapping critical bands, called as critical bandwidth. Each critical bandwidth represents one bark in the Bark scale [2]. The 24 center frequencies used in Bark scale are fixed. Bark scale is based on "loudness" measurement of speech signal [3]. Maximum sampling frequency suggested for Bark scale is 15,500 Hz. For frequencies higher than above values, downsampling of the speech signal needs to be done which leads to aliasing effect [4]. To avoid this effect, cutoff frequencies are formulated based on diatonic scale. The proposed diatonic frequency cepstral coefficients (DFCC) algorithm allows to extend the range of sampling frequencies to higher values above 15,500 Hz. Some speech recognition applications require sampling frequencies higher than 15,500 Hz like sampling frequency used for recording music is 44,000 Hz. A diatonic scale is based on the "quality" of sound. The proposed DFCC algorithm improved the accuracy and precision of classification over existing methods considerably. The feature vector produced by diatonic scale is used for classification of spoken words using feed forward neural network (NN). NN provides prediction accuracy which can be misleading if prediction values of the different classes are almost equal. As a result, the network misclassifies the pattern to the highest prediction accuracy class though it may not be the correct class. Hence, fuzzification can be used at the output of NN [5]. Application of fuzzy inference system at the output of NN tries to overcome abovementioned problem by reclassifying only the confusing patterns using proposed hybrid classifier. The overall accuracy as well as precision of the entire system is improved using hybrid neuro fuzzy (NF) classifier. The paper is organized as follows: review of related work in Sect. 2, methodology developed by authors in Sect. 3, results and discussion in Sect. 4 followed by conclusion in Sect. 5.

2 Related Work

Isolated word speech recognition in native languages spoken in different parts of India is an ongoing topic of research. An audio corpus of Marathi language containing 28,420 isolated words and 17,470 sentences is prepared by Santosh Gaikwad et al. It is available for experimentation in speech recognition system [6]. Recognition of Marathi numerals has been effectively implemented using Mel frequency cepstral coefficients (MFCC) and dynamic time warping (DTW) [7]. Speaker independent Urdu speech recognition system for district names of Pakistan has been studied and analyzed. The result proved that the system performs better if it is accent independent [8]. DNN–HMM hybrid model is developed for Chinese–English mix database to promote multilingual research [9]. The maximum word error rate achieved for Chinese database is 48.38%. Neuro fuzzy classification approach is used for recognizing sign language for hearing and speech impaired people. Combination of neural network and fuzzy logic resulted in 78% classification accuracy for the system [10].

Neuro fuzzy quantification is used for customer emotional state recognition. Multiple adaptive neuro fuzzy inference system with 95.29% classification accuracy is been developed [11]. Gujarati language speech corpora is prepared for speech recognition system. HMM is used as a classification algorithm. The system gives 95.9% word recognition rate [12]. Hierarchical speech recognition system is developed for Hindi language. The system aims at phoneme recognition [13]. MFCC and linear predictive cepstral coefficients techniques are used with HMM-based classification for Kannada language isolated spoken words [14]. The developed system was evaluated with 79% accuracy. Speech recognition is done for mobile applications in Tamil language using MFCC and dynamic time warping template matching [15]. Comparative study of different speech recognition techniques has been done by Gaikwad et al. One can choose an appropriate feature extraction technique with its merits and demerits [16]. Spoken Arabic digit recognition is done using MFCC and DTW techniques [17]. An average accuracy of 87% without normalization of feature vector was obtained.

3 Methodology

A novel algorithm to generate diatonic frequency cepstral coefficients using a musical scale called as diatonic scale is proposed in this work. The scale is based on quality of sound. Quality distinguishes two sounds of same pitch and loudness. Sound produced by musical instruments consists of series of tones with different frequencies called as overtones. Fundamental tone and series of overtones are in the ratio 1:2:3 called as harmonics. A diatonic scale has a series of eight notes between starting note and last note in an octave. It uses 8 frequencies starting on the tonic fundamental frequency, $F0$, and ending one octave above that at $2F0$. The scale then repeats, with 8 frequencies from $2F0$ to $4F0$, each being exactly twice the frequency considered as base octave. The next octave uses frequencies exactly four times those in the base octave. Input to the diatonic scale is the fundamental frequency $F0$ based on what the remaining octave frequencies are calculated. This property of the diatonic scale is used to generate cutoff frequencies in the feature extraction stage. Table 1 gives the ratio with $F0$ in the first octave span as well as the interval between consecutive notes. Human ear is most sensitive to frequencies between 300 Hz and 3 kHz. Hence, first octave starting with fundamental frequency 300 Hz is repeated to get diatonic scale frequencies within abovementioned range. Figure 1 shows the diatonic scale curve plotted with number of diatonic frequencies on x scale and corresponding frequency value in Hz on y scale.

Table 1 First octave used in diatonic scale

Ratio with $F0$	1	9/8	5/4	4/3	3/2	5/3	15/8	2
Interval	9/8	10/9	16/15	9/8	10/9	9/8	16/15	–

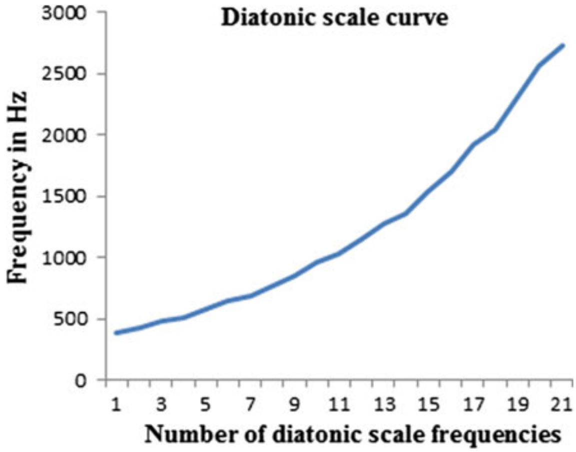


Fig. 1 Diatonic scale curve

The process of speech recognition starting from input speech to output recognized word is illustrated in Fig. 2. In this section, feature extraction from the recorded signal, training, and speech recognition stages is described in detail.

3.1 Feature Extraction Using DFCC

Speaker utters a word which is recorded through microphone. This speech signal recorded has voiced part, unvoiced part, and silence portion. The voiced portion of the signal is needed to be separated out for further processing. Preprocessing of the speech signal includes voiced part detection and preemphasis.

Voiced Part Detection Let y be the input speech signal with length L and sampling frequency f_s.

$$y = \{y_0, y_1, y_2, \ldots, y_{n-1}\} \quad (1)$$

Speech processing requires nonstationary speech signal to be divided into number of frames denoted as N_f. When speech signals with different length undergo framing, unequal number of frames is generated. To maintain constant length of feature vector generated from these signals, zero padding is done to the frames. This affects the patterns generated by feature extraction algorithm reducing the recognition rate. To

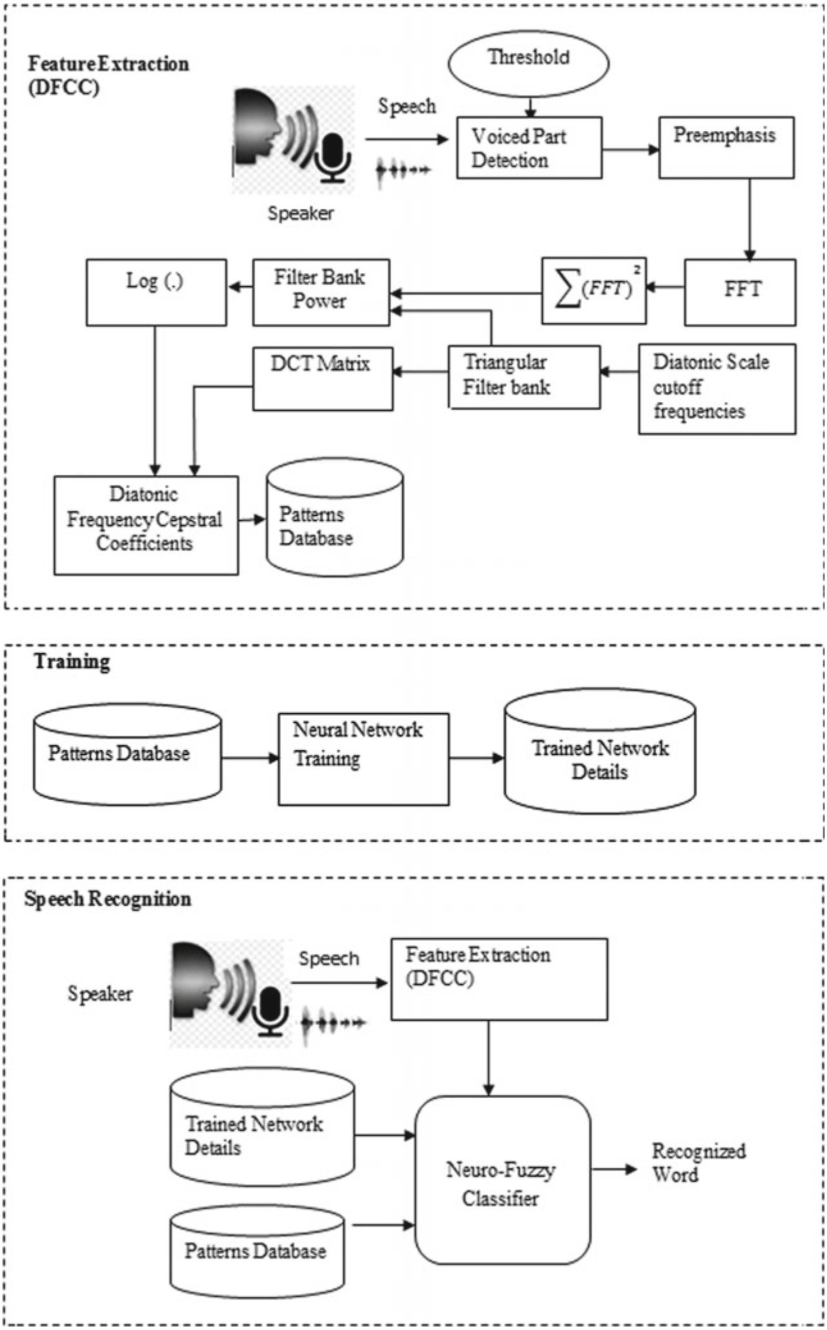


Fig. 2 Speech recognition using DFCC feature extraction and hybrid NF classifier

overcome this problem, the proposed algorithm uses variable length frames where the frame length depends on the length of the original signal.

Irrespective of the length of the recorded signal, equal number of frames is generated using Eq. (2) in the proposed work.

$$N_{\mathrm{f}} = \frac{(L - l)}{\mathrm{step}} + 1, \quad l = 0.02 f_{\mathrm{s}}, \mathrm{step} = 0.01 f_{\mathrm{s}} \tag{2}$$

Distortions are added due to number of edges created by frames. Hanning window is applied to each frame to reduce the effect of distortions.

$$w(i) = 0.5\left(1 - \cos\left(\frac{2\pi i}{l - 1}\right)\right) \quad 1 \leq i \leq l \tag{3}$$

Since short time energy (STE) of the voiced part of a speech signal is high [18], we calculate STE per frame.

$$\mathrm{STE}_l = \sum_{i=1}^{l} (y_i * w_i)^2 \tag{4}$$

If STE is greater than a threshold value, it is considered to be the voiced part of the speech sample.

$$y_1(n) = y(n), \quad \mathrm{STE}_l \geq \lambda \tag{5}$$

where $y_1(n)$ is the voiced part of the signal.

Preemphasis Preemphasis filter is applied to the speech signal to spectrally flatten the signal. This filtering helps to keep high-frequency components between the samples n and $n - 1$. Equation (6) represents the preemphasis of speech signal $y_1(n)$. The value of α is chosen as 0.97, $y_2(n)$ is preemphasized speech signal.

$$y_2(n) = y_1(n) - \alpha(y_1(n - 1)) \tag{6}$$

Diatonic scale cutoff frequencies To generate the cutoff frequencies, diatonic scale octave is repeated till maximum available frequency of the speech signal. Since the sampling frequency used in the proposed work is 16 kHz, we need to repeat the diatonic octave till maximum frequency of 8000 Hz. Table 2 shows 31 cutoff frequencies thus obtained using diatonic scale. The ratio of frequencies with $F0$ is repeated as shown in Table 1 for each consecutive octave.

Table 2 Diatonic scale cut off frequencies in Hz

384	426.66	480	512	572	640	682.66	768
853.33	960	1024	1152	1280	1365.33	1536	1706.66
1920	2048	2304	2560	2730.66	3072	3413.33	3840
4096	4608	5120	5461.33	6144	6826.66	7680	

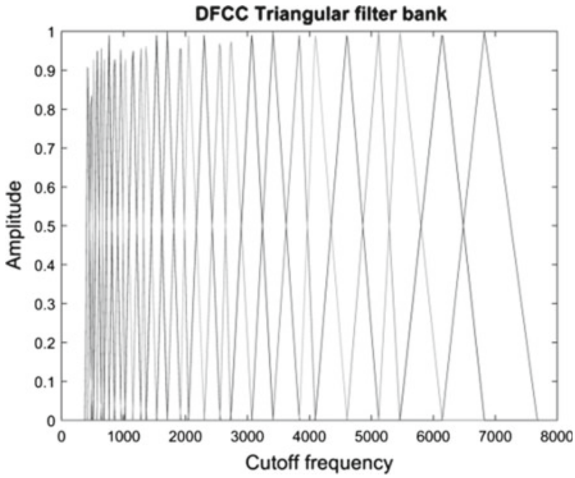


Fig. 3 DFCC triangular filter bank

Next step in the generation of filterbank is calculating upslope and downslope coefficients for the filter bank channels. M is the number of filterbank channels. The value of M used in the proposed work is 29.

Upslope coefficients

$$k = f \geq \mathrm{cf}(m) \text{ and } f \leq \mathrm{cf}(m+1), \quad 1 \leq m \leq M$$
$$H(m,k) = \frac{(f(k) - \mathrm{cf}(m))}{(\mathrm{cf}(m+1) - \mathrm{cf}(m))}, \quad 1 \leq k \leq K \tag{7}$$

Downslope coefficients

$$k = f \geq \mathrm{cf}(m+1) \text{ and } f \leq \mathrm{cf}(m+2), \quad 1 \leq m \leq M$$
$$H(m,k) = \frac{(\mathrm{cf}(m+2) - f(k))}{(\mathrm{cf}(m+2) - \mathrm{cf}(m+1))}, \quad 1 \leq k \leq K \tag{8}$$

where cf is the cutoff frequency for each triangular filter bank, $K = n\mathrm{FFT}/2$

Figure 3 displays the triangular filter bank generated using diatonic frequencies. It represents 29 filter banks placed between 300 and 8000 Hz. Discrete Fourier transform $F(u)$ with length N is calculated for each frame. Here $y_{2\mathrm{frame}}$ is the preemphasized speech signal in a frame.

$$F(u) = \sum_{z=0}^{N} y_{2\text{frame}}(z) \mathrm{e}^{\frac{-j2\pi z u}{N}} \tag{9}$$

Magnitude spectrum

$$|F(u)| = \sqrt{\text{real}(F(u))^2 + \text{imag}(F(u))^2} \tag{10}$$

Power spectrum

$$P(u) = |F(u)|^2 \tag{11}$$

Triangular filter bank $H(M, K)$ is applied to power spectrum to generate filter bank power vector fbp(M).

$$\text{fbp}(M) = \sum_{k=1}^{k} H(M, k) P(k, 1), \quad K = n\text{FFT}/2 \tag{12}$$

A discrete cosine transform matrix is created for each filter bank with number of cepstral coefficients as one of the dimensions.

$$\begin{aligned} &\text{dctm}(p, m) = w(p) \sum_{m=1}^{M} \cos\left(\frac{\pi}{2M}(2m-1)(p-1)\right), \quad 1 \le p \le n\text{cep} \\ &\text{where} \quad w(p) = \begin{cases} \frac{1}{\sqrt{N}} & p = 1 \\ \sqrt{\frac{2}{N}} & 2 \le p \le M \end{cases} \\ &n\text{cep} = \text{number of cepstral coefficients} \end{aligned} \tag{13}$$

Filter bank power matrix is applied to discrete cosine transform matrix (dct_m) to get cepstral coefficients vector, CC.

$$\begin{aligned} &\text{CC}(p, j) = \sum_{m=1}^{M} (\text{dctm}(p, m) l\text{fbp}(m, j)) \\ &\text{where } 1 \le p \le n\text{cep}, \quad 1 \le j \le N_\text{f}, \\ &l\text{fbp} = \ln(\text{fbp}) \end{aligned} \tag{14}$$

N_f is the number of frames in which the speech signal is divided. The diatonic scale cepstral coefficients vector "CC" generated in Eq. 14 is stored in a patterns database which is input to the next training stage.

3.2 Training

The feature vector from the feature extraction stage is used for training neural network. It is a supervised training model that simulates the complex behavior of human brain. Training yields the class identification of each sample where we know the number of desired classes. It calculates the matching scores between the input pattern and sample pattern of a class. The proposed algorithm uses NN training. Test bed used for training NN and classification is recorded by us and named as Spoken Marathi Numeral Dataset (SMND) [19]. It consists of 7500 utterances of Marathi numerals from 0 ("Shunya") to 9 ("Nau"). 75 Speakers in the age group 18–55 with different gender and dialect pronounced the digits in succession. Recording was done in Mono mode with professional microphone and stored in.wav file format. Sampling frequency used for recording is 16 kHz. Closed room with provisions to minimize reverberation of sound was used for recording. 50% of these are used for training the network. The number of input neurons is 120. The number of hidden neurons used is 27 and the number of output neurons is 10 depending upon desired 10 classes at the output representing 10 numerals available in SMND. At the output of this stage, trained network details with best matching scores are obtained.

3.3 Speech Recognition

A hybrid classifier is used in this stage. This is a two-step process. The information retrieved from trained network details is used to determine desired output and calculate the error. This stage selects the class with best matching score and generates output which is closest to the input pattern. Input to this stage is the patterns database generated from the feature extraction stage and trained network details generated by training stage. Standalone NN when used for classification of isolated words produces prediction variables with almost similar values. In such cases, if the prediction variable of the wrong class is larger with minor difference than the actual class, NN produces wrong output. To overcome this problem, the prediction variables produced by NN are passed to fuzzy controller. The job of fuzzy controller is to improve the recognition accuracy by correctly classifying above patterns. Fuzzy logic is a part of fuzzy inference system that integrates human decision making in the form of IF-THEN rules [20]. Combination of fuzzy sets that represent some sort of uncertainty and IF-THEN rules determine the correct output class. Figure 4 represents the NF classifier architecture. This hybrid system has NN in the first step. NN classifier has inputs I_1 to $I_{m,}$ input neurons a_1 to $a_{m,}$ hidden neurons h_1 to h_n and NN output classes c_1 to c_k. The NN output is further given as input to fuzzy controller. The fuzzy inference system provides patterns to be reclassified to the next k nearest neighbors (k-NN) classifier. k-NN is a simple machine learning algorithm that uses a distance function to classify a new case. k-NN generates the correct classified word at the output. The role of fuzzy controller is to decide which misclassified

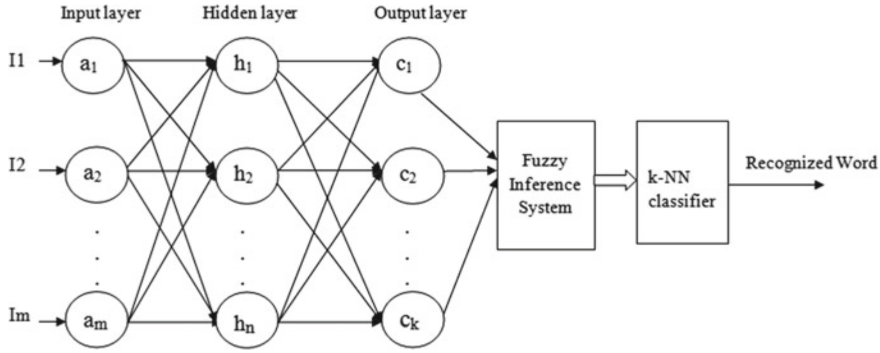


Fig. 4 Hybrid NF classifier architecture

patterns in the output of NN have very close prediction variables. These will be sent to next classifier. This improves the recognition rate of the system. It reduces overall computational complexity as few selected patterns are sent for reclassification. The proposed algorithm uses a triangular membership function in the fuzzy controller in order to define fuzzy rules using linguistic variables. This step is necessary to decide whether a pattern has to be sent to next classifier. We have used Euclidean distance measure as a similarity measure in *k*-NN classifier. *k*-NN classifies the pattern into correct class in majority cases. The proposed two-step hybrid classifier including NF as the first step followed by *k*-NN as the second step improves accuracy as well as precision of classification significantly over standalone NN classifier.

4 Results and Discussion

NN training of the patterns generated from SMND database is done using proposed DFCC feature extraction method. The dataset has utterances of ten Marathi digits from 0 to 9. It has 750 samples of each digit contributing to 7500 samples. Out of these 50% samples are used for training and 50% are used for testing of classifier. Thus, number of training and testing samples is 3750 each. Performance of the proposed DFCC method with an existing feature extraction method MFCC is compared. Table 3 represents recognition accuracy of ten digits for MFCC and DFCC using NN classifier. DFCC shows significant improvement over existing MFCC algorithm for all the ten digits. Figure 5 shows the plot of the recognition accuracy for MFCC and DFCC. From the results plotted, it can be concluded that the accuracy of digits 4, 5, 6, and 7 is lesser as compared to remaining digits. To improve overall accuracy, number of true positives for these digits should be increased which is achieved by implementing a hybrid classifier. We consider a case of digit 1 misclassified as digit 3. The prediction values at the output of NN classifier for the two digits are 0.4697 and 0.5729, respectively. NN misclassifies such patterns where the difference between

Table 3 Classification accuracy in % for ten digits using NN classifier

Digit category	0	1	2	3	4	5	6	7	8	9	Avg.
MFCC	96.3	94.8	93.6	96.8	93.3	92.3	92.2	88.7	91.1	96.4	93.5
DFCC	97.1	95.8	97.8	96.3	93.7	93.0	92.3	90.0	98.1	97.3	95.2

Table 4 Classification accuracy of DFCC algorithm for different datasets using NF classifier

Database	Accuracy (%)	Precision (%)
SMND	99.23	96.19
FSDD	99.00	95.56

Table 5 Summary of experimental results

Feature extraction	Accuracy (%)		
	SVM	NN	NF
MFCC	89.66	93.59	94.63
DFCC	89.77	95.15	99.23

prediction variables is less. Fuzzy controller assigns linguistic labels to these values as "High" as per triangular membership function. Since both of them have same linguistic label, these patterns are sent to k-NN classifier. In step 2 of the classification, k-NN classifies these patterns into actual classes. Figure 6 represents the comparative results of NN classifier and proposed NF classifier. It represents improvement by NF classifier in the number of true positives. True positives are the number of correct classified digits. From the figure, we can conclude that the proposed NF classifier improves classification accuracy for all the ten digits. We have tested the proposed method on another database comprising of English numerals. The database is available online for downloading known as free spoken digits dataset (FSDD). It has 2000 utterances of ten digits from 0 to 9 recorded by 4 speakers. We have used 1400 samples for training and 600 samples for testing purpose. Table 4 shows the comparative results for both the datasets. DFCC is used for feature extraction and classification is done with NF classifier. Accuracy and precision values are comparable for both the databases. Accuracy is the ratio of number of correct predictions and total number of patterns in the dataset. Precision is the ratio of number of correct positive predictions to the total number of positive predictions. Thus, the proposed algorithms can recognize words from different languages effectively. Table 5 shows the summary of the experimental results. MFCC and DFCC feature extraction techniques are compared for different classifiers. Classifiers used are NN, support vector machine (SVM) and proposed NF classifier. SVM is an existing supervised learning algorithm used for comparative analysis of the proposed hybrid classifier in our experimentation. The results are compared on the basis of average classification accuracy values. From the summary of results, we can conclude that both the proposed algorithms DFCC feature extraction and hybrid classifier outperform the existing techniques.

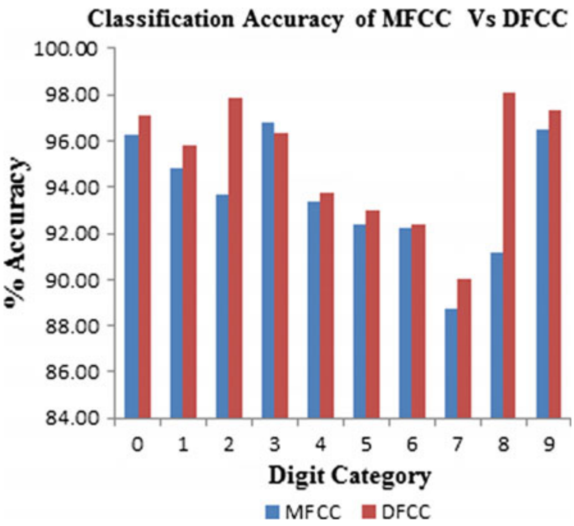


Fig. 5 Comparison of MFCC and DFCC classification accuracy using NN classifier

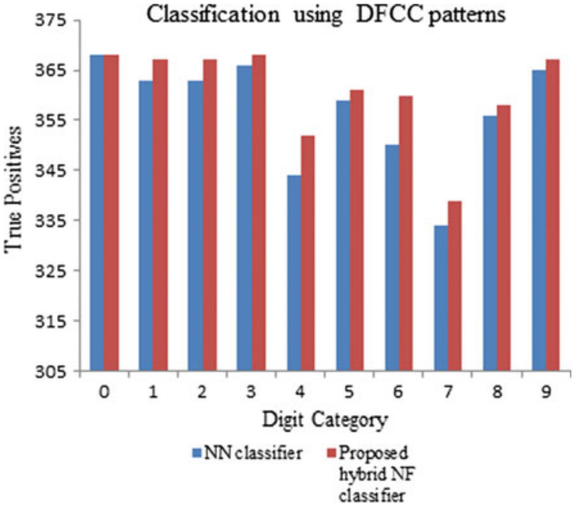


Fig. 6 Comparison of the number of true positives generated by NN classifier and proposed NF classifier using DFCC feature extraction

5 Conclusion

We have developed a language independent isolated word speech recognition system. The proposed system contributes at two stages. First, we have proposed a feature extraction method DFCC based on diatonic scale frequencies. This algorithm gives an overall classification accuracy 95.20% whereas existing MFCC technique has 93.50% accuracy. DFCC pattern generation can be used for a number of applications like speech recognition, music classification since diatonic scale cut off frequencies can be generated for any sampling frequency value. In the second stage of proposed work, we have designed a hybrid NF classifier to recognize patterns generated by DFCC algorithm. The combination of DFCC along with NF classifier generates an accuracy 99.23% which is significantly higher than existing SVM and NN classification methods. The NF classifier has a novel and unique implementation in the proposed work. This implementation ensures improved recognition rate for the speech recognition system irrespective of gender, age, and dialect variations of speakers.

References

1. Gupta D, Bansal P, Choudhary K (2018) The state of the art of feature extraction techniques in speech recognition. In: Agrawal S, Devi A, Wason R, Bansal P (eds) Speech and language processing for human-machine communications, vol 664. Advances in intelligent systems and computing. Springer, Singapore, pp 197–207
2. Lin Y, Abdulla WH (2015) Principles of psychoacoustics. Audio watermark. Springer, Cham, pp 15–49
3. Shanon BJ, Paliwal KK (2003) A comparative study of filter bank spacing for speech recognition. In: Microelectronic engineering research conference, Brisbane, pp 1–3
4. Hsieh SH, Lu CS, Pei SC (2013) Sparse fast fourier transform by downsampling. In: IEEE International conference on acoustics, Vancouver, pp 5637–5641
5. Bhavsar H, Trivedi J (2018) Image based sign language recognition using neuro fuzzy approach. Int J Sci Res Comput Sci, Eng Inform Technol, IJSRCSEIT 3:487–491
6. Gaikwad S, Gawali B, Mehrotra S (2013) Creation of Marathi speech corpus for automatic speech recognition. In: Conference on Asian spoken language research and evaluation (O-COCOSDA/CASLRE), Gurgaon, pp 1–5
7. Gedam YK, Magare SS, Dabhade AC, Deshmukh RR (2014) Development of automatic speech recognition of Marathi numerals. Int J Eng Innovative Technol (IJEIT) 3:198–203
8. Qasim M, Nawaz S, Hussain S, Habib T (2016) Urdu speech recognition system for district names of Pakistan. In: Conference of the oriental chapter of international committee for coordination and standardization of speech databases and assessment technique, Bali, pp 28–32
9. Wang D, Tang Z, Tang D, Chen Q (2016) A Chinese-English Mixlingual database and a speech recognition baseline. In: Conference of the oriental chapter of international committee for coordination and standardization of speech databases and assessment technique, Bali, pp 84–88
10. Li W, Hu X, Gravina R, Fortino G (2017) A neuro-fuzzy fatigue tracking and classification system for wheelchair users. IEEE Access 5:19420–19431
11. Diago L, Kitaoka T, Hagiwara I, Kambayashi T (2011) Neuro-fuzzy quantification of personal perceptions of facial images based on a limited dataset. IEEE Trans Neural Networks 22:2422–2432
12. Tailor JH, Shah DB (2018) HMM based light weight speech recognition system for gujarati language. In: Mishra D, Nayak M, Joshi A (eds) Information and communication technology for sustainable development. Lecture notes in networks and systems, vol 10. Springer, Singapore
13. Samudravijaya K, Ahuja R, Bondale N, Jose T, Krishnan S, Poddar P, Raveendran R (1998) A feature based hierarchical speech recognition system for Hindi. Sadhana. 23:313–340
14. Sneha V, Hardhika G, JeevaPriya K, Gupta D (2018) Isolated Kannada speech recognition using HTK-A detailed approach. In: Saeed K, Chaki N, Pati B, Bakshi S, Mohapatra D (eds) Process in advanced computing and intelligent engineering. Advances in intelligent systems and computing, vol 564. Springer, Singapore
15. Dalmiya CP, Dharun VS, Rajesh KP, (2013) An efficient method for tamil speech recognition using MFCC and DTW mobile applications. In: IEEE conference on information and communication technologies, Jeju Island, pp 1263–1268

16. Gaikwad S, Gawali B, Yannawar P (2010) A review on speech recognition technique. Int J Comput App 3:16–24
17. Ganoun A, Almerhag I (2012) Performance analysis of spoken arabic digits recognition techniques. J Electron Sci Technol 10:153–157
18. Jalil M, Butt FA, Malik A (2013) Short time energy, magnitude, zero crossing rate and autocorrelation measurement for discriminating voiced and unvoiced segments of speech signals. In: The international conference on technological advances in electrical, electronics and computer engineering (TAEECE), Konya, pp 208–212
19. Kondhalkar H, Mukherji P (2017) A database of Marathi numerals for speech data mining. Int J Adv Res Sci Eng 6:395–399
20. Bai Y, Wang D (2006) Fundamentals of fuzzy logic control-fuzzy sets, fuzzy rules and defuzzifications. In: Bai Y, Zhuang H, Wang D (eds) Advanced fuzzy logic technologies in industrial applications, advances in industrial control. Springer, London, pp 17–36

Performance Analysis of Fuzzy Rough Assisted Classification and Segmentation of Paper ECG Using Mutual Information and Dependency Metric

Archana Ratnaparkhi, Dattatraya Bormane and Rajesh Ghongade

Abstract The paper aims at the development of fuzzy rough set-based dimensionality reduction for discrimination of electrocardiogram into six classes. ECG acquired by the offline method is in the form of coloured strips. Morphological features are estimated using eigenvalues of Hessian matrix in order to enhance the characteristic points, which are seen as peaks in ECG images. Binarization of the image is carried out using a threshold that maximizes entropy for appropriate extraction of the fiducial features from the background. Various image processing algorithms enhance the image which is utilized for feature extraction. The dataset produced comprises the feature vector consisting of 79 features and 1 decision class for 6 classes of ECG. Extensive analysis of dimensionality reduction has been done to have relevant and nonredundant attributes. Fuzzy rough domain has been explored to take into account the extreme variability and vagueness in the ECG. Optimal feature set is subjected to fuzzification using Gaussian membership function. Further, fuzzy rough set concepts help in defining a consistent rule set to obtain the appropriate decision class. Classification accuracy of unfuzzified dataset is compared with the fuzzified dataset. Semantics of the data are well preserved using fuzzy rough sets and are seen from the performance metrics like accuracy, sensitivity and specificity. The proposed model is named as Fuzzy Rough ECG Image Classifier (FREIC) which can be deployed easily for clinical use as well as experimental use.

A. Ratnaparkhi (✉)
AISSMS-IOIT, SPPU, Pune, India
e-mail: archana.ratnaparkhi@gmail.com

D. Bormane
AISSMS COE, SPPU, Pune, India

R. Ghongade
BVCOE, Bharti Vidyapeeth, Pune, India

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_77

1 Introduction

Arrhythmias cause a change in the normal rhythm of the heart which when sustained for a long time may result in irreparable damage to the heart as they actually cause a change in the normal rhythm of heart rate [1]. Thousands of beats are recorded by the Holter device carried out by the patients and immediate detection of the abnormality from such a huge database after preprocessing becomes a herculean task. 12-lead ECG system provides the information on complete ECG health from 12 different views. The ECG rhythm strips record the ECG which is analysed by the physician. Advancement in the field of Biomedical Image Processing has greatly assisted in feature extraction and dimensionality reduction that immensely contributes in quick and faster assessment of the patient. Dimensionality reduction to achieve appropriate feature extraction is generally a fundamental step in data processing. Removal of irrelevant features or redundant features enhances the discriminatory or predictive power of the algorithm. Selection of number of attributes which effectively constitute to increase the classification accuracy is possible due to rough set-based feature extraction. Rough set theory deals with datasets which are discrete in nature. Continuous-valued attributes need to undergo discretization which might add to information loss and further loss in classification accuracy. The previous approaches in ECG preprocessing and delineation has been widely done using Pan-Tompkins algorithm. However, the temporal information in ECG seems to get distorted by the use of filtering techniques like bandpass filtering [2], Kalman filtering and ensemble averaging [3] filtering [4]. Delineation of ECG has been done using various approaches like envelope detection and slope criteria for various segments in ECG that is seen in [5, 6]. The second-order differentiation and wavelet transform for extraction of the characteristic points is done by Duangsoithong and Windeatt [7] and Chan [8]. One can find the application of artificial neural networks for ECG delineation and classification [9]. Adaptive filtering and dynamic time warping has also been worked upon by Thakor and Zhu [3] and Laguna et al. [4]. Discrete wavelet transform has been widely used for the detection of QRS complex and other features [10]. For detecting congenital long QT syndrome DWT has been effectively used by Chevalier et al. [11]. Hybrid neural network approach in classifying the ECG into several classes has been done by Dokur and Olmez [12] using Daubechies wavelet (db2) [13]. The paper proposes a fuzzy rough assisted ECG classification of the ECG images that are obtained from the ECG rhythm strips.

1.1 Preliminaries

1.1.1 Rough Set Theory

Rough sets theory, a mathematical tool for representing datasets in an approximate manner, has been a primary tool to deal with imprecise and incomplete information.

It was originally proposed by Professor Pawlak in 1982 [1]. The original theory was put forth in such a simplified manner that it encouraged the researchers to apply this theory in varied fields ranging from image processing to higher dimensional signal processing systems and intelligent fault diagnosis systems [5]. Basically, the Rough Set Theory (RST) comprises of generation of reduced set of attributes called reducts which are used to form rules that classify a test dataset. A unique advantage of RST is that it deals with data and needs no other information about the dataset. Fuzzy set theory, rough set theory and probability theory have been an attraction for researchers by virtue of their immense capability to deal with uncertainty in the information. The structure of rough sets with abstract algebra is done in [8–10]. Rough sets have been effectively combined with soft computing tools in [5–10, 13, 14]. The study of evolutionary algorithms combined with RST has been done. Particle swarm organization, genetic algorithms and ant colony optimizations have been used in [15]. Thus, RST mainly focuses on attribute reduction and rule generation to be further processed by intelligent systems. However, as per the literature, RST suffers due to the crisp representation of the data. An acceptable range of variation in the input values is allowed in fuzzy domain. Hence with the advent of Fuzzy rough set as done by Jensen and Shen [15], representation of the biosignals has greatly simplified.

1.1.2 Fuzzy Rough Set Theory

A hybrid fuzzy rough set model first proposed by Dubois and Prade [16] was later extended and/or modified by many authors, and was applied successfully in various domains, most notably machine learning [17]. The upper and lower approximation sets in rough set theory have been fuzzified in fuzzy rough domain. However, certain principles as suggested by Jensen and Cornelis [17] need to be followed. And they have been stated as follows The set A may be generalized to a fuzzy set in X, allowing that objects can belong to a given concept to varying degrees. Instead of assessing objects indiscernibility, we may measure their approximate equality. As a result, objects are categorized into classes, or granules, with soft boundaries based on their similarity to one another. As such, abrupt transitions between classes are replaced by gradual ones, allowing that an element can belong (to varying degrees) to more than one class. The concept of belongingness to more than one class is able to efficiently handle the variability in ECG. The next section explains the use of fuzzy rough set theory applied to ECG feature vector which is extracted from the preprocessed ECG image (Fig. 1).

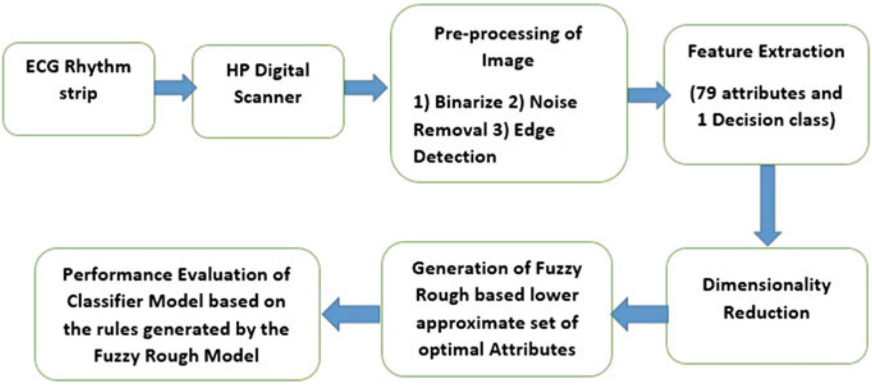


Fig. 1 Methodology

1.2 Methodology

1.2.1 Acquisition and Preprocessing

As given in the figure, ECG rhythm strip as recorded by the 12-lead ECG acquisition system is considered for further processing. This strip is scanned by HP digital scanner to get a good quality and clear image. Such a scanned image is stored in TIFF format and preprocessed using ImageJ toolbox. The preprocessing basically involves binarizing the image to enhance the ECG morphology. Further noise filtering is done using Gaussian filter which not only filters the image but also smooths it.

1.2.2 Feature Extraction and Dimensionality Reduction

The filtered binary image is subjected to various image processing tools that enhance the ECG morphology. The Sobel Edge detector highlights the edges in the ECG which helps in better feature vector formation. The features extracted include various parameters of the Hessian matrix. Second-order local image intensity variations around the selected pixel are described by the Hessian matrix. An orthonormal coordinate system is generated due to eigenvector decomposition which is perfectly aligned with the second-order structure of the image. A feature vector consisting of 79 attributes is further subjected to dimensionality reduction. Various standard attribute evaluators that are used in the literature have been tested.

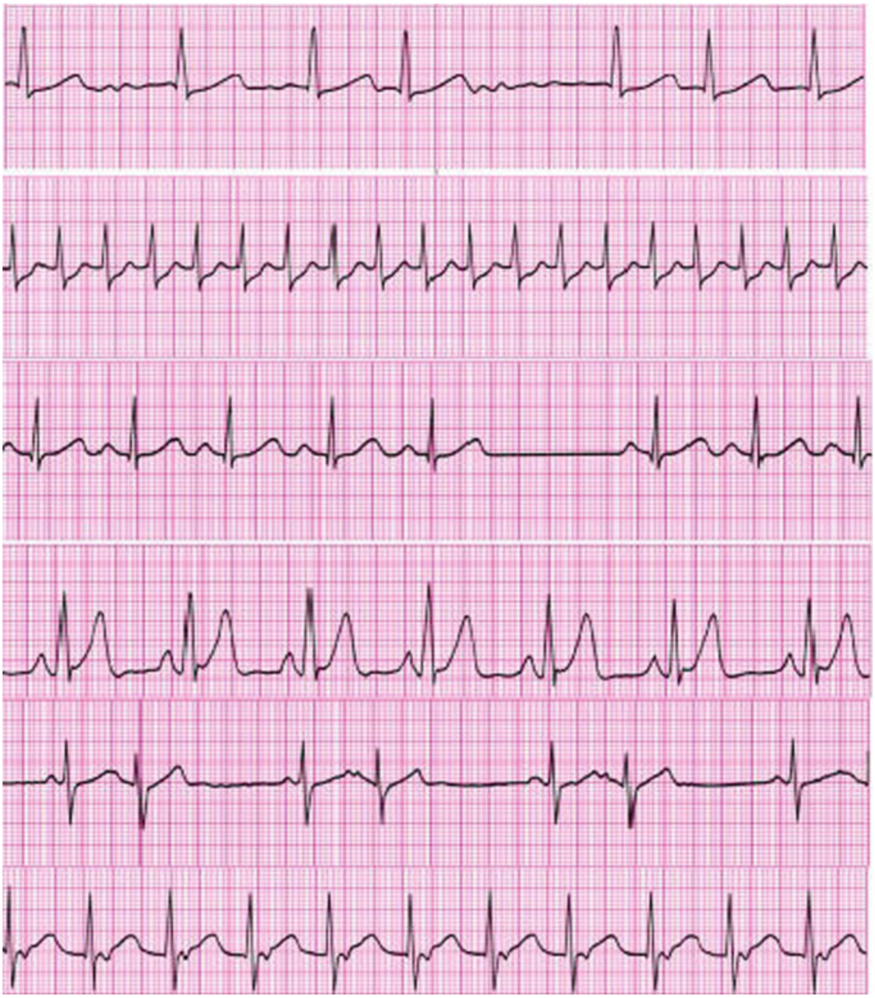


Fig. 2 ECG rhythm strip

1.3 Fuzzy Rough-Based Rule Generation and Classifier Evaluation

The optimal attributes generated are used to generate fuzzy rough-based lower approximate sets. Fuzzification of the inputs helps in incorporating the uncertainty, vagueness and extreme variability in the ECG. The novelty in Fuzzy Rough ECG Image Classifier (FREIC) lies in the use of fuzzy rough set-based rule induction using optimal attributes. According to [10], the QUICK-REDUCT algorithm is as follows:

$Quickreduct(C, D)$

$Input1$: C, the set of conditional attributes;

$Input2$: D,the set of decision features

$Output$: R,the feature subset

1.R $\leftarrow$ { }

2.while $\gamma_R(D) \neq \gamma_C(D)$

3.$T \leftarrow R$

4.for each x $\in$ (C - R)

$if\ \gamma_{R\cup x}(D) > \gamma_T(D)$

$T \leftarrow R \cup \{x\}$

5.$R \leftarrow T$

6. $\mathrm{Re}\,turn \quad R$

The above fuzzy rough attributes evaluator helps in finding the optimal attribute set. The proposed FREIC classifier is compared with standard classifiers. Performance metrics for the same are discussed in the next section.

1.4 Experimentation and Results

The ECG strips obtained from 12-lead ECG acquisition system is scanned using the digital scanner. The figures below show the images obtained after scanning. These strips show six classes of Arrhythmia. The binary ECG image enhances the ECG morphology and assists in appropriate edge detection which is seen in the next Fig. 3. These preprocessed images are used for feature extraction using ImageJ toolbox. Hessian of the image is taken which sharply highlights the characteristic points in ECG. 79 such attributes further are subjected to Dimensionality Reduction. The dimensionality reduction using fuzzy rough metric is comparable to the reduction obtained using correlation-based method. The proposed model FREIC based on fuzzy rough metric looks for attributes which are relevant and nonredundant. Such attributes are then evaluated on the basis of their dependency metric (dependency of the decision class on a certain attribute). Thus, eventually dual dimensionality reduction occurs. The classification accuracy obtained using tenfold cross-validation on standard clas-

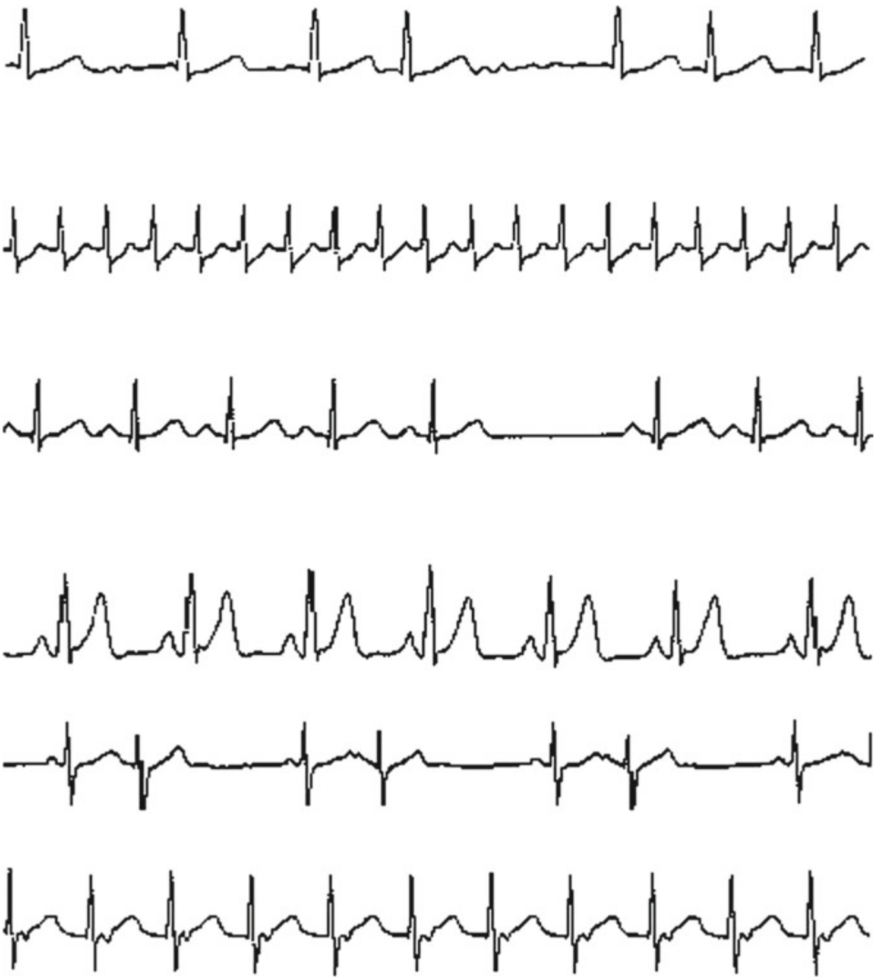


Fig. 3 Binary image of the ECG strip

sifiers like SVM, MLP, Random Forest, Rough sets and Fuzzy classifier is shown in the graph below. Comprehensive comparison of the classifiers is done. The proposed Fuzzy Rough ECG Image classifier reduces the number of attributes to nine which is considerably low as compared to SVM or MLP-based classification. Comparative analysis of the accuracies clearly shows the efficacy of rough sets and fuzzy rough sets-based classifier as they outperform the traditional methods in terms of optimal attribute selection as well as accuracy as is seen in ROC curves (Figs. 2, 4, 5, 6, 7, 8, 9 and 10).

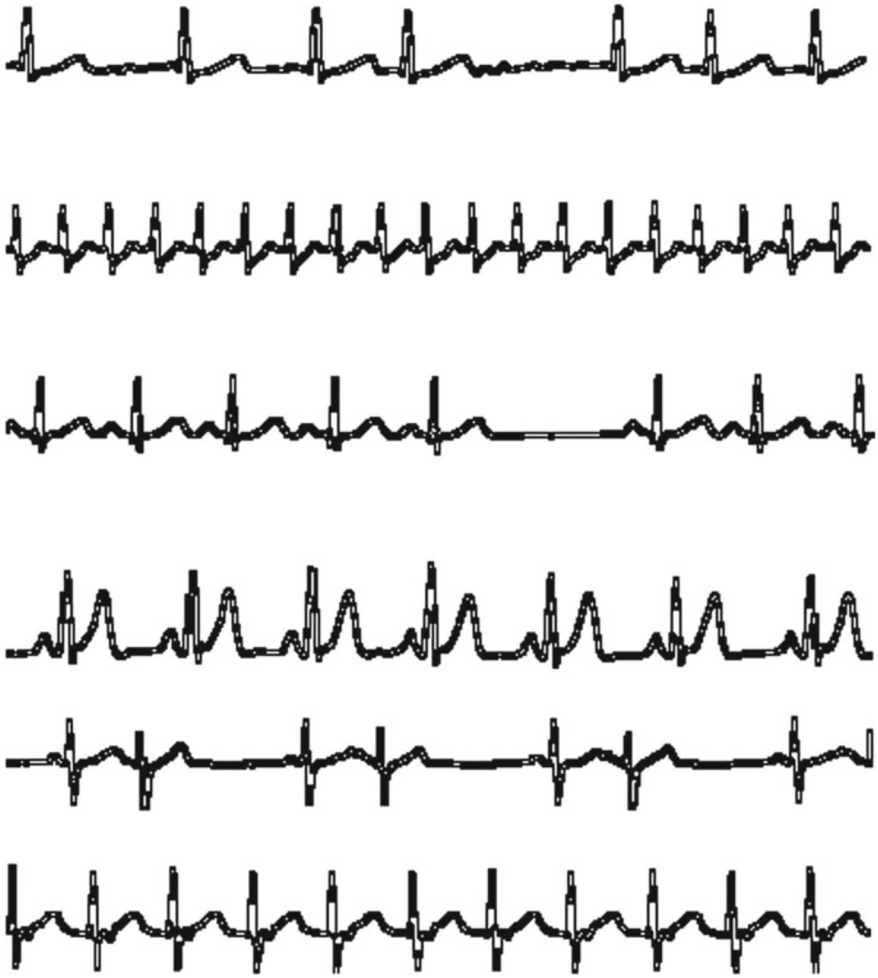


Fig. 4 Binary image of the edge detected ECG strip

1.5 Conclusion

The proposed Fuzzy Rough ECG Image classifier exploits the concepts of lower and upper approximate sets from fuzzy rough set theory for ECG rhythm strips which is completely novel. The algorithm not only outperforms other traditional classifiers in terms of accuracy by a comfortable margin, but also competes with much-proven correlation-based attribute evaluators to achieve an optimal attribute set. Additional resilience has been achieved by the image processing tools that effectively enhance the morphology of the ECG signal and assist in achieving high accuracy.

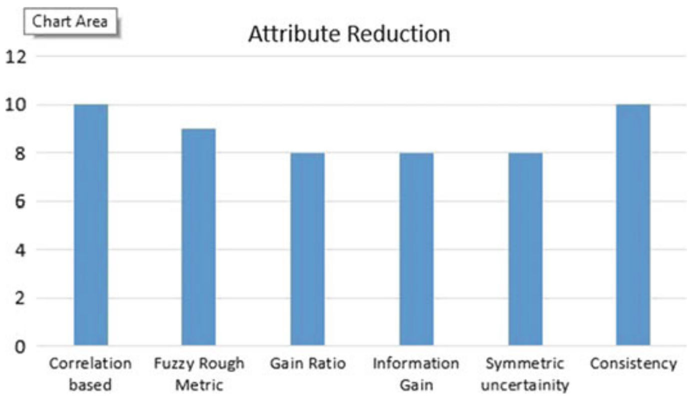


Fig. 5 Attribute reduction using standard attribute evaluators

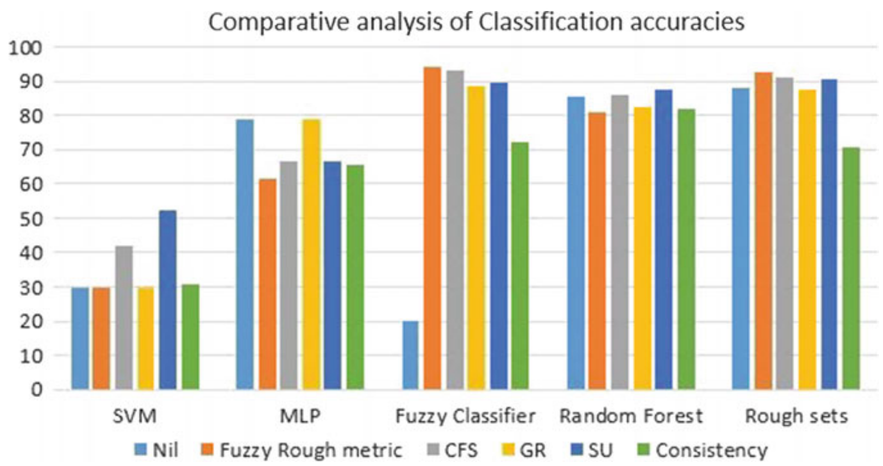


Fig. 6 Comparative analysis of classification accuracies

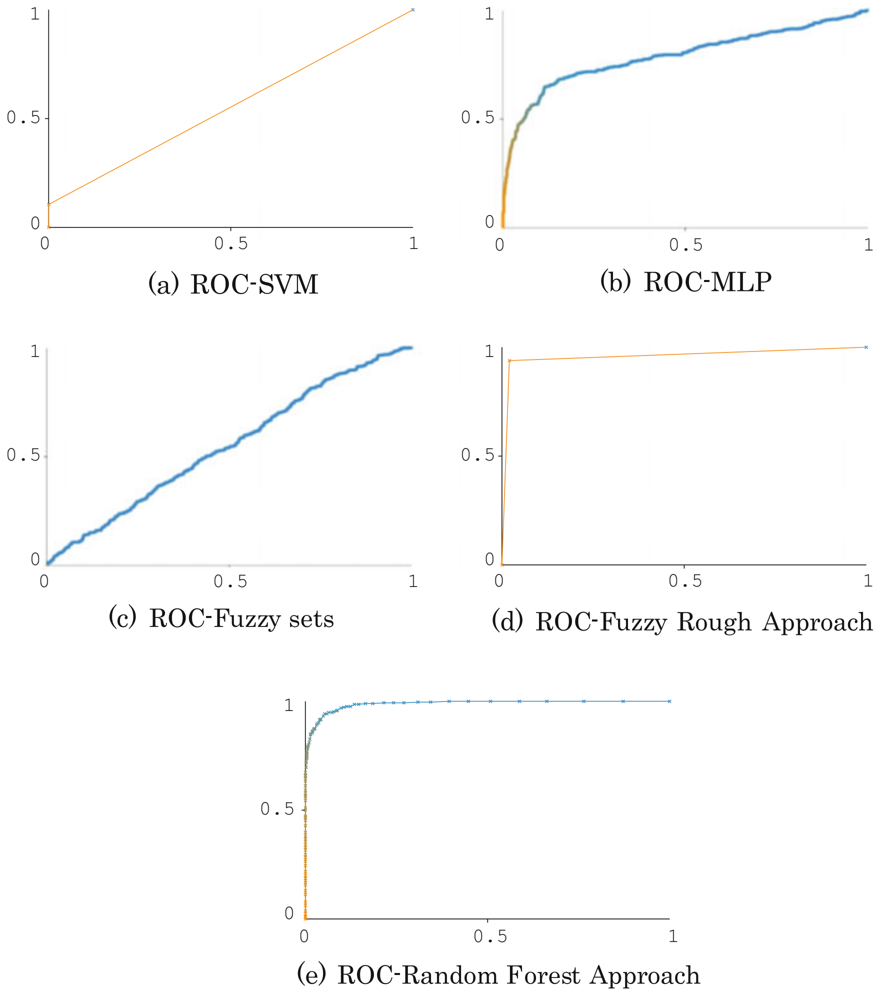


Fig. 7 ROC curves: classifiers

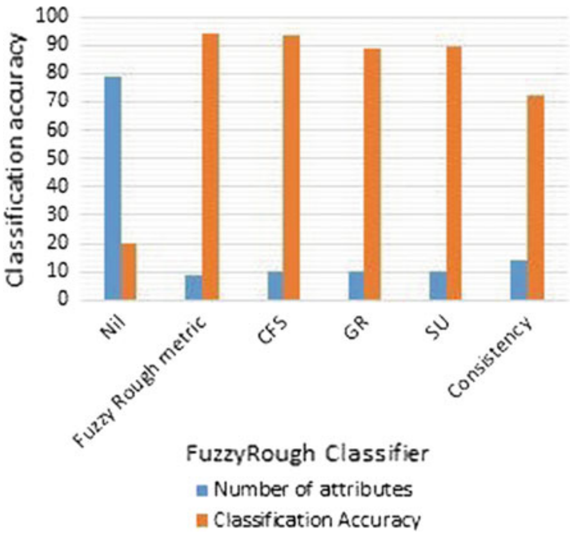


Fig. 8 SVM

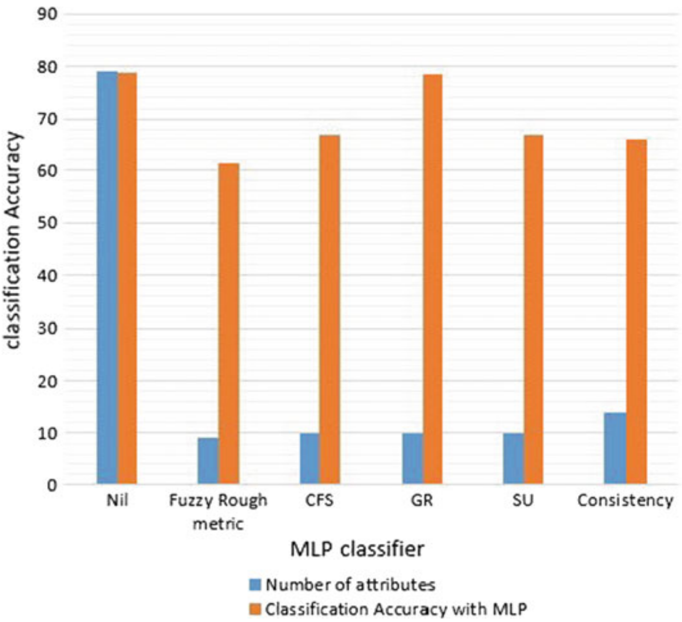


Fig. 9 MLP

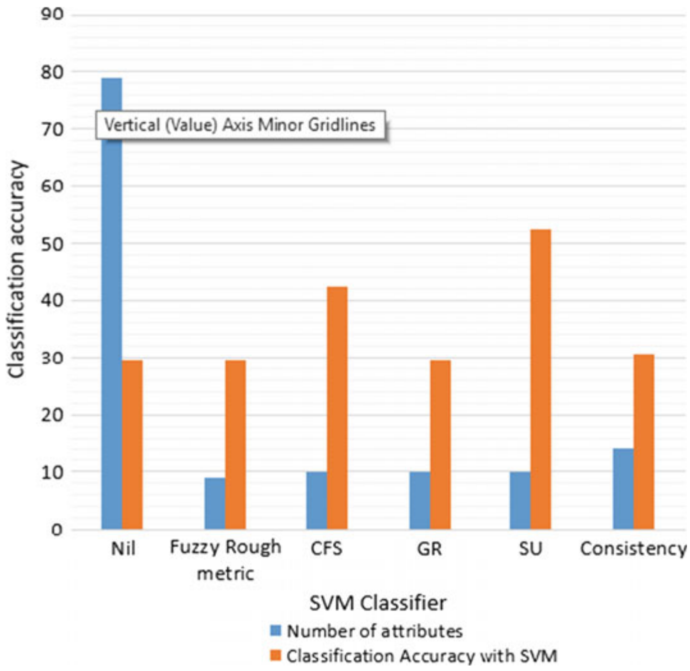


Fig. 10 Fuzzy rough sets

References

1. Tan WW, Foo CL, Chua TW (2007) Type-2 fuzzy system for ecg arrhythmic classification. In: IEEE International Fuzzy systems conference, 2007. FUZZ-IEEE 2007. IEEE, pp 1–6
2. Christov II, Daskalov IK (1999) Filtering of electromyogram artifacts from the electrocardiogram. Med Eng Phys 21(10):731–736
3. Thakor NV, Zhu Y-S (1991) Applications of adaptive filtering to ecg analysis: noise cancellation and arrhythmia detection. IEEE Trans Biomed Eng 38(8):785–794
4. Laguna P, Jané R, Meste O, Poon PW, Caminal P, Rix H, Thakor NV (1992) Adaptive filter for event-related bioelectric signals using an impulse correlated reference input: comparison with signal averaging techniques. IEEE Trans Biomed Eng 39(10):1032–1044
5. Zhang Q, Xie Q, Wang G (2016) A survey on rough set theory and its applications. CAAI Trans Intell Technol 1
6. Polkowski L (2002) Topology of rough sets. In Rough Sets. Springer, Berlin, pp 331–360
7. Duangsoithong R, Windeatt T (2011) Hybrid correlation and causal feature selection for ensemble classifiers. In: Ensembles in machine learning applications. Springer, Berlin, pp 97–115
8. Chan YT (2012) Wavelet basics. Springer Science & Business Media
9. Ghongade R et al (2007) A brief performance evaluation of ecg feature extraction techniques for artificial neural network based classification. In TENCON 2007-2007 IEEE region 10 conference. IEEE, pp 1–4
10. Martínez JP, Almeida R, Olmos S, Rocha AP, Laguna P (2004) A wavelet-based ecg delineator: evaluation on standard databases. IEEE Trans Biomed Eng 51(4):570–581
11. Chevalier P et al (2001) Non-invasive testing of acquired long QT syndrome: evidence for multiple arrhythmogenic substrates. Cardiovasc Res 50(2):386–398

12. Dokur Z, Ölmez T (2001) ECG beat classification by a novel hybrid neural network. Comput Methods Programs Biomed 66(2–3):167–181
13. Addison PS (2005) Wavelet transforms and the ecg: a review. Physiol Measur 26(5):R155
14. Mitra S, Mitra M, Chaudhuri BB (2006) A rough-set-based inference engine for ecg classification. IEEE Trans Instrum Measur 55(6):2198–2206
15. Jensen R, Shen Q (2007) Fuzzy-rough sets assisted attribute selection. IEEE Trans Fuzzy Syst 15(1):73–89
16. Dubois D, Prade H (1990) Rough fuzzy sets and fuzzy rough sets. Int J Gen Syst 17(2–3):191–209
17. Jensen R, Cornelis C (2011) Fuzzy-rough nearest neighbour classification and prediction. Theoret Comput Sci 412(42):5871–5884

A Mixed Reality Workspace Using Telepresence System

Viken Parikh and Mansi Khara

Abstract Even after years of advancement from telephone technology to video conferencing, it eventually does not provide the environment of real human interaction. It has been a challenge ever since. The system we propose here would provide an environment in the user's reality, where the user can interact with the person on the other side of the conversation with the help of a wearable mixed reality headset. The user will get a feeling of coexistence with objects and people who are far away. This advancement in technology will lead to a world, where people can connect with others who are only virtually present. This paper will present the technical system for teleporting virtual objects and virtual participants in the real-world environment. It will present a system where 3D audio and real-time communication will help build a more realistic environment with the presence of virtual participants and objects in 3D space.

1 Introduction

Telepresence is the hypothetical movement of material objects from one place to another without physically traversing the distance between them. By merging Mixed Reality and Telepresence, we can get a real world experience with virtual objects where people can communicate and interact with each other without being physically present. In this era, traveling seems to be more and more difficult, telepresence is an emerging research area and companies are moving forward to make this possible. It will seem astonishing to see how people travel across the globe virtually. This will save both time and money and will prove to be the next generation of technology. In this paper, we provide an approach to regenerate objects in a 3D space and display it to a user using light-field technology in Mixed Reality headset. The objects will

V. Parikh (✉) · M. Khara
K. J. Somaiya College of Engineering, Vidyavihar, Mumbai, India
e-mail: viken.parikh@somaiya.edu

M. Khara
e-mail: mansi.khara@somaiya.edu

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_78

be captured with the help of 8 trinocular cameras where the RGB images along with the depths of these images will be captured, and the 3D model will be reconstructed. The color RGB image will provide color to the 3D model such that it appears to be the same as the original object. Finally, the object will be displayed on the other side to the person wearing the headset. The applications of the proposed model range from Virtual Meetings, where people can conduct important meetings virtually and can meet their loved ones from across the globe. This can also be useful in Virtual Tourism, where one can roam the world by sitting at just one place and capture memories, by storing the virtual meetings and regenerating them when needed. This can also be used in the field of Gaming, where playing a game will feel seamless, for example, the player playing the battle actually on the battlefield.

The paper is organized as follows, Sect. 2 shows the basic terminologies used by us throughout the paper. Section 3 gives a broader view of similar work done in the past, Sect. 4 describes the actual methodology used, applications are described under Sect. 5, Sect. 6 concludes our paper, and Sect. 7 described the work that can be continued in the future.

2 Basic Terminologies Used

2.1 Telepresence

It is the use of virtual reality technology for remote control of machinery or for apparent participation in distant events.

2.2 NIR

Near-infrared, a region of the infrared spectrum of light used for near-infrared spectroscopy. It is used for multi-depth mapping in imaging. For more details, refer [1].

2.3 Light-Field Technology for Imaging

Conventional images record the projection of light rays on a 2D plane as compared to light fields, which describe the distribution of light rays in a free space, consisting of their position, angle, and radiance. Light-field technology uses this idea to capture these image data for application in mixed reality. For more details, refer [2].

2.4 Binaural Recording

It is a method of recording sound using two microphones, arranged with the intent to create a 3D stereo sound sensation for the listener. It will perceive the listener of actually being in the room with the performers or instruments.

3 Related Work

The major differences and advantages what MR brings as compared to AR and VR are that, in AR, the position of the objects displayed to the user is not relative to the user's position in the 3D space. The user will be displayed data on their screen as it is, on top of the real-world environment. While VR creates a separate digital space, where the user can only interact with 3D objects and data that have been digitally made. MR provides the advantage over both of these technologies by providing relative positioning of objects on top of the real-world environment with the help of smart consumer glasses or displays.

There has been a lot of research work in this field. But the current research lacks the development of a system that can be used for teleconferencing with the help of mixed reality in real-time communication with efficient results.

HoloLens provides an AR technology but lacks the light-field technology to enable Mixed Reality. Many of the VR headsets have some related health problems like dizziness and headache due to low frame rates. They also lack the efficiency to place virtual world objects in the real world. Locomotion is restricted by the presence of limited object manipulation option in 3D space.

These technical smart video headsets and glasses are highly expensive and require continuous high bandwidth network usage and processing power. Hence compression, low-power processing, and immediate transmission of data are a few of the objectives while designing a mixed reality system.

For telepresence of human models, the need for continuous tracking is required at a high frame rate to provide a continuous communication experience to the participants.

There have been attempts but, none being able to demonstrate a mixed reality workspace and hence the need for research on this topic.

4 Methodology

Our main aim here is to create a two-way interaction between two participants in mixed reality. This project will allow a local participant and a distant participant to interact with each other and also interact with other virtual objects in a 3D space. The main importance here is a low latency model with the highest possible quality of the regenerated objects in 3D space.

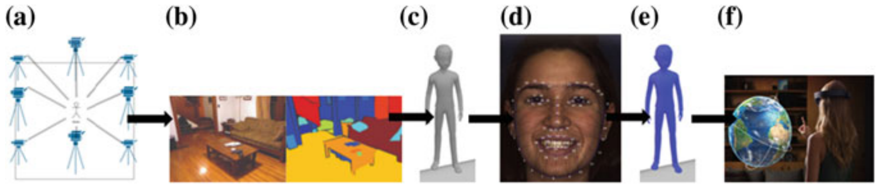


Fig. 1 **a** Capturing image. **b** Depth capturing and preprocessing. **c** 3D reconstruction. **d** Face and body tracking [8]. **e** Merging the colors with the 3D model. **f** Display in mixed reality to the remote participant

The model shown in Fig. 1 is a quick overview of how our proposed model would work. This overview provides a workflow of how the physical setup will enable interaction with virtual objects in the real-world environment.

4.1 Capturing Images

The participant and user will be in a room with a setup of 8 capturing trinocular cameras. The 8 trinocular cameras will be arranged as shown in Fig. 2a. The edge cameras are arranged higher as compared to the corner cameras. This would help create a better 3D model while reconstruction with the help of 8 images from 8 different positions. Each trinocular camera will be arranged as shown in Fig. 2b, which will consist of two Near-Infrared (NIR) cameras and one color camera.

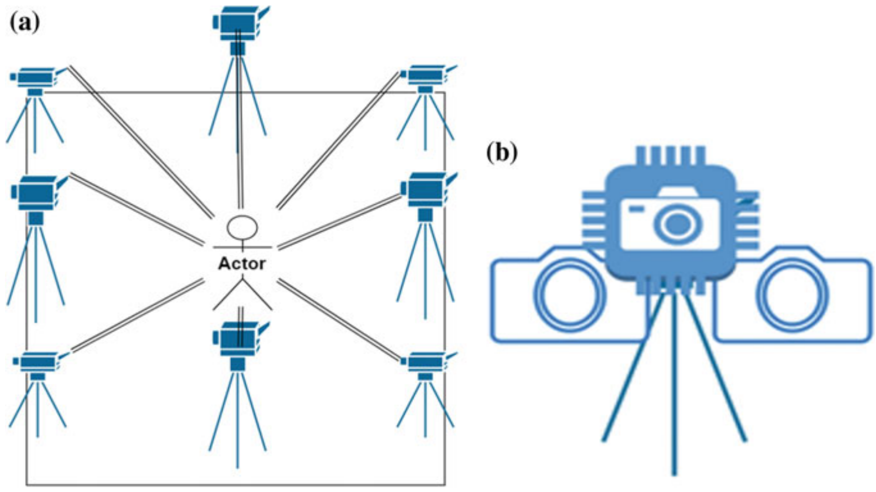


Fig. 2 **a** Arrangement of 8 trinocular cameras. **b** Trinocular camera setup with 2 NIR cameras and 1 color image capturing camera

- The distance of the object and position in the real world will be estimated with the help of the Depth cameras [3]. The first image captured is an RGB image and the second image is an NIR Image. There are several options available to capture depth images. We would propose Orbbec Astra Mini camera, which can be used for NIR image capturing. Kinect can also be used.
- For the RGB images, high-quality digital cameras will be used.
- All the cameras would be synchronized running at 60 fps.

4.2 Creating Virtual Objects in 3D Space

Figure 3 demonstrates how a 3D object can be placed along with real-world objects in the virtual space.

We use Autodesk Maya application to create 3D objects. The libraries already have a rich collection of redeveloped 3D objects. Other tools like Blender, 3ds Max can also be used. We use these objects and import them into Unity3d. Unity3d provides an environment which can be replicated to be similar to that of the real-world environment. Now the objects created can be placed in the 3D space of Unity3d, which would eventually be placed in the same virtual space as that of the models of human participants. We will use object recognition to detect a type of object, and based on object pairs created, we will place the objects in the real-world space. The light and placements of the objects can be trained to be placed in the places where it is most likely to be based on object placements pairs. For example, if an object is a human model, it is most likely to be placed on a chair, sofa, or floor based on his position and posture. A fruit basket object is most likely to be placed on a table.

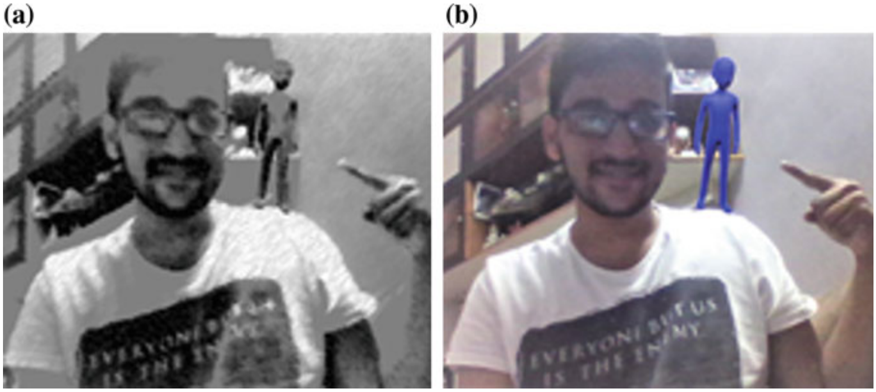


Fig. 3 **a** Image of virtual object silhouette to be added on top of the real world using mixed reality display. **b** Finally recreated 3D object added to the mixed reality display. Viken Parikh, K J Somaiya College of Engineering

4.3 Generating Silhouettes to Create a 3D Model of Real-World Objects

With the help of 2D images, we create silhouettes. These silhouettes help in image compression and also help us to extract the areas of interest like human models from the images. With the help of silhouettes, we can get the actual shape of the object, and a reverse algorithm can help us reconstruct that shape [4]. As we have multiple images, it will help us to create a better 3D model of the object.

4.4 3D Model Reconstruction

With the help of segmentation from the silhouettes, we get the objects to be reconstructed for our 3D model. Using temporal consistency, we can separate and create a more consistent model of our object. A statistical function can be used which contains objects of the same class with different appearances to be applied on the geometrical object captured from the real-world data. This mapping of the 3D object with common appearance statistics will provide more room for compression and better accuracy for low latency platforms (Fig. 4).

- Color Texturing or Matting
  After fusing depth data, a polygonal 3D model is extracted from its implicit volumetric representation (TSDF) with marching cubes, and then, this model is textured using the 8-input RGB images [5]. The approach presented for color texturing can classify unobserved views which form ghostly artifacts with traditional approaches.
- Human and Face tracking
  There are two types of human tracking, whole-body tracking and face tracking. Whole-body tracking corresponds to tracking a general tracking rather than small facial movements or gestures. Figure 5b shows whole-body tracking.

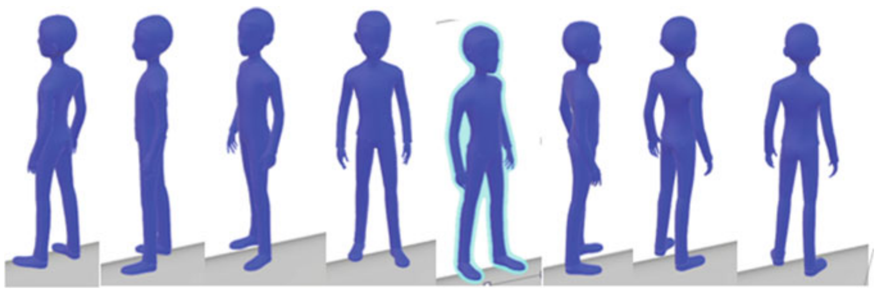


Fig. 4 3D object created used 8 different images from 8 different angles captured

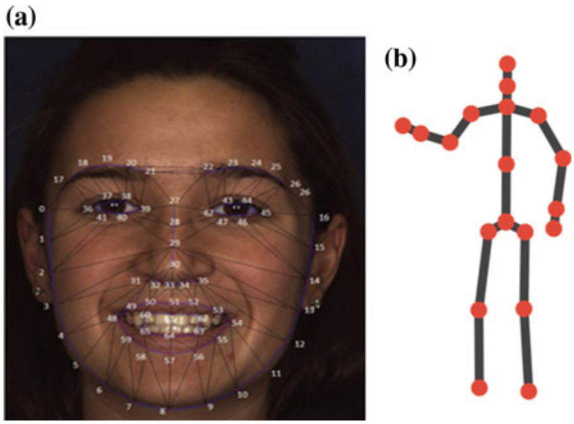


Fig. 5 **a** Face tracking. **b** Whole-body tracking [8]

On the other hand, facial tracking is dedicated to tracking these small behaviors. For objects with minute details like face, a posture which changes quickly with respect to time, we apply face tracking for rigid and nonrigid facial tracking. Figure 5a shows face tracking. For more read [6].

4.5 3D Audio

Until a few years back, audio was recorded by using two methods: mono and stereo. Mono uses a single microphone to pick up a sound, while stereo uses two, spaced apart from each other.

The binaural audio setup is a simulation of how we receive sound in real life. That creates an experience, a user would if they were actually there.

Binaural recording is similar to the stereo method of recording. It records by placing two microphones in ear-like cavities on either side of a stand or dummy head. Though similar, it is better than stereo, as the microphones are placed on a person's ears, or inside ear-like imitations on a dummy's head. As the dummy head recreates the density and shape of a human head, these microphones capture and process sound exactly as it would be heard by human ears to preserve interaural cues.

In order for the user to have a complete experience of the real-world environment, along with the virtual one, the user cannot wear a headset or earpieces. Therefore, the speakers of the 3D audio would be placed just about the ears in order for a user to have a complete virtual and real-world experience (Fig. 6).

Fig. 6 A binaural audio recording device for capturing 3D audio

4.6 Compression

Compression is important for rendering such huge data to a distant location. There are several ways of performing this compression, some of them include mesh compression and point cloud compression.

- Mesh Geometry
  Mesh compression was used to compress the polygonal model before sending over the Internet by serializing vertices. Two approaches were tested, 3Dzip for transmission over TCP and FFmpeg for UDP [7].

4.7 Processing and Transmission

- Parallelization across CPU or GPU
  First is the location of the reconstruction algorithm. In the model-based approach, it is executed on the CPU and in the render-based approach, on the GPU. This is because the former contains much branching, whereas, the latter is undertaking the same parallel instructions on different data [7].
- Transmission over the internet
  Based on the accuracy required by the user and bandwidth available, the number of participants, part of the mixed reality telepresence system can vary. For a system that can provide for a high bandwidth, it can support the telepresence for 5–10 users in the same reality simultaneously.

4.8 Display in Mixed Reality

- Light-Field Technology
  Light Field is the term given to light rays traveling in all the directions in a given volume of space. Unlike conventional images, which record the 2D projection of the light rays by angularly integrating the rays at each pixel and missing out on important information, a light field describes the distribution of light rays in free space. This allows capturing richer information from the world. This is done using a micro-lens array, which is placed between the main lens and the camera sensor. Light-Field Technology solves one of the biggest challenges, which is enabling virtual objects to look real from both far and near distances.
- Gesture-Controlled Mixed reality headsets
  We propose a mixed reality headset that uses light-field technology to display 3D objects in the real world [2]. This will allow users to adjust the focus of near and far objects and have a better experience of the real world. The headsets will display 3D objects, which look very similar to real-world objects with low latency.

With the advancements of light-field technology, the MR headset will be able to adjust the position of the object with respect to the real-world environment. The user will have a provision to change the position of the object in 3D space with the help of gesture patterns. By default, the objects will be placed in 3D space based on object pairs.

These holograms of 3D objects on the top of real-world environments enable the user to have an experience of being with multiple users along with an ability to manipulate their positions and interact with them.

5 Applications

There are several other applications of teleportation. Some of them include the following:

- A new innovation in the field of Telecommunications
  From the traditional voice calls to video calls, technology is becoming more and more advanced. The concept of telepresence will give birth to people being virtually present at some other place, in 3D. Figure 7a is an example of the same.
- Gaming
  While gaming is still restricted to 2D or 3D with limited functionality, using Virtual Reality. Telepresence can be a whole new experience with actual objects inside the game can be seen in your surroundings, different players can play the game together by being virtually present. This innovation in teleportation will surely make gaming a wonderful experience.

- Creating memories
  Traditionally, the only way to store memories, be it some good holiday, or some important event was with the help of taking photographs. Telepresence will help us store the memories in 3D, resize them, replay them whenever one wishes to.
- Virtual Tourism

People can teleport virtually from their place to a tourist location. This will save their time as well as money. Figure 7b shows a woman traveling to a distant location from her home by wearing a mixed reality headset.

6 Conclusion

Thus, we present a powerful object reconstruction technique, which can be used to map 3D objects in real time and present the same at a different location. The process needs high-quality cameras, mixed reality wearable headsets, and software tools to map and reconstruct the objects. We have mentioned the potential applications of our model and how this technology can enhance the future generation.

7 Future Work

New approaches can be considered, for capturing better quality pictures and reducing the latencies to regenerate them at the destination. A large number of users can be accommodated per mixed reality teleconferencing workspace by either increasing the bandwidth available, increasing the processing power and by improving the compression algorithm. Further, low-cost mixed reality displays like Aryzon can be made with additional features like 3D audio and gesture control so that a larger group of users can benefit from it.

Fig. 7 **a** People sitting in a conference virtually. **b** A woman enjoying a tourist location with the help of telepresence

Acknowledgements The paper, its research, and its content is ethically ours and any of the other content used has been correctly referenced to its source. This is to clarify the ethical compliance of our paper to be published in Springer.

References

1. Ng R (2006) Digital light field photography
2. https://www.technologyreview.com/s/610458/vr-is-still-a-novelty-but-googles-light-field-technology-could-make-it-serious-art/
3. Xiong Y (2014) Automatic 3D human modeling: an initial stage towards 2-way inside interaction in mixed reality
4. Speakman S (2013) Dynamic pattern detection with temporal consistency and connectivity constraints
5. Orts-Escolano S et al (2016) Holoportation: virtual 3D teleportation in real-time
6. Mangold K, Shaw JA, Vollmer M (2013) The physics of near-infrared photography. Eur J Phys 34. IOP Publishing Ltd
7. Roberts DJ et al (2015) withyou—an experimental end-to-end telepresence system using video-based reconstruction. IEEE J Sel Top Signal Process 9
8. https://www.cl.cam.ac.uk/research/rainbow/projects/clmz/labels/indices.jpg

Detection of Static and Dynamic Abnormal Activities in Crowded Areas Using Hybrid Clustering

M. R. Sumalatha, P. Lakshmi Harika, J. Aravind, S. Dhaarani and P. Rajavi

Abstract In Computer Vision, to monitor the activities, behavior, and other changing information, surveillance is used. Surveillance is used by government organizations, private companies for security purposes. In video processing, anomaly is generally considered as a rarely occurring event. In a crowded area, it is impossible to monitor the occasionally moving objects and each person's behavior. The main objective is to design a framework that detects the occasionally moving objects and abnormal human activities in the video. Histogram of Oriented Gradient feature extraction is used for the detection of an occasionally moving object and abnormality detection involves in computing the motion map by the flow of motion vectors in a scene that detects the change in movement. The experimental analysis demonstrates the effectiveness of this approach which is efficient to run in real time achieves 96% performance, however, for effective validation of the system is tested with standard UMN datasets and own datasets.

1 Introduction

In current world video surveillance plays a vital role. It is a process of monitoring video clips for the behavior, activities, or other suspicious information for the purpose of managing or protecting the society. In a congested area, it is impossible to monitor each person's behavior. The behavior of each person varies. Some behave normally, while few behave abnormally. Therefore, it is a tedious process to monitor suspicious behavior. Arousing, fighting, rioting, jaywalkers, wrong way, etc., is considered as some of the suspicious behavior. Surveillance system has become an important aspect of security and a necessity to keep proper checking. With this fast growing technological world, the use of video surveillance is also increased. Now

M. R. Sumalatha (✉) · P. Lakshmi Harika (✉) · J. Aravind · S. Dhaarani · P. Rajavi
Department of Information Technology, MIT Campus, Anna University, Chennai, India
e-mail: sumalatha.ramachandran@gmail.com

P. Lakshmi Harika
e-mail: lakshmi.harika558@gmail.com

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_79

a day, video surveillance is used in almost all areas. The main aim is designing the model to detect the abnormality in the human behavior either static or dynamic and to recognize the crowd event in the video sequence based on features extracted and the movement of motion vectors. Human anomaly detection in crowded areas is a great challenge. The existing framework has certain limitations like slow processing time, needs high storage space, cannot find occluded objects, need previous results as input.

2 Related Works

Lin et al. [1] designed a framework for unattended object detection. Immobile foreground is extracted and it is used to detect the unattended objects by backtracking the owner trajectories.

Lin et al. [2] framed a method which used grab cut segmentation [3] for unattended object detection. Background for detection is developed using dual Gaussian Mixture learning model [4]. But the work is limited to the complex environments and mainly focused on increasing efficiency.

Jadhav et al. [5] devised an algorithm for unattended object detection. By subtracting the frame from the background, the objects in the foreground are obtained. Then, using the feature of objects like shape, height, color, and size the unattended objects are identified and classified.

Jaouedi et al. [6] proposed a method to recognize the human actions by inspecting their behavior. In their work, they tracked the moving humans by using Kalman filtering [7]. Then their actions are determined by using their motion vectors and K nearest neighbor method is used for classification.

Kumar et al. [8] developed an approach for recognition of the activities of human in video sequence. This is done by finding the optical flow [9] of the edges of humans. But their work is based only on the known data set.

Zhu et al. [10] framed an approach to human behavior analysis by using the hybrid model. In this the behavior is inspected by extracting global silhouette feature and optical flow method is used to estimate their motions. These are the limitations in existing survey and it has been overcome in the proposed system.

3 Proposed System

The proposed system detects occasionally moving objects and uncommon human activities in video frames. The architecture for detection of occasionally moving objects and uncommon human activities is given in Fig. 1.

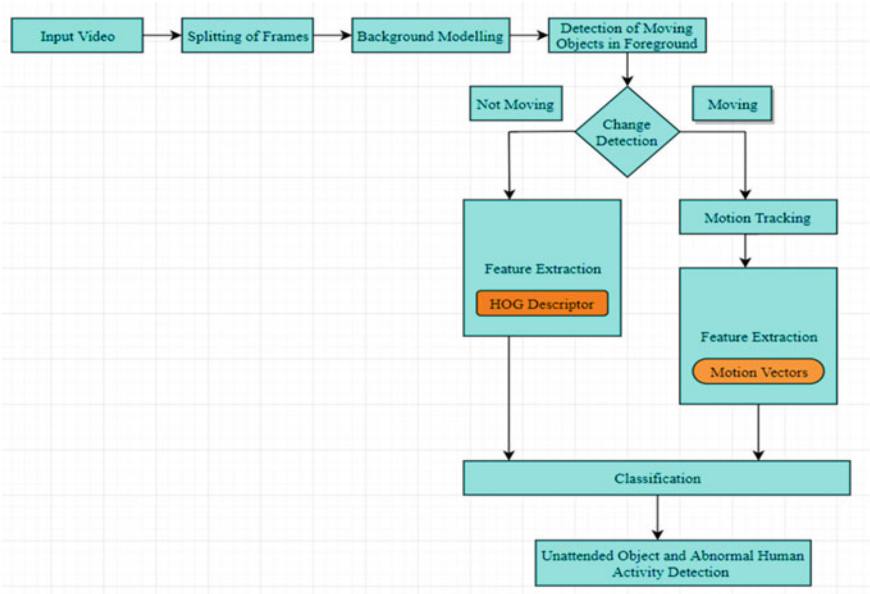


Fig. 1 Architecture for activity detection system

3.1 Preprocessing

The input to the system is a video file. The input video is split into frames and they are processed in a successive manner. Frames are sequences of images. The input video is preprocessed to remove noise and for the conversion of RGB frame to gray-scale frame. The detection of occasionally moving objects and uncommon human activities is done by comparing the frames of video with background frame.

3.2 Background Modeling

Modeling of background is used to obtain accurate background from input video file. Gaussian Mixture model in Eq. (1) is used to model the background and the background is generated with first few frames.

Gaussian distribution

$$B = \arg\min\left(\sum_{K=1}^{b} w_k > T\right), \tag{1}$$

where B is a background model, T is a threshold, w_k is a weight factor of pixels.

Initially, video frames are given as an input to the system. A background model is initialized for all pixels in the background frame. The value of pixels in consecutive frames is checked in a successive manner to know whether they are pixels of background.

3.3 Occasionally Moving Static Object Detection

In video frames, subtraction between the images is a straightforward approach to detect the changes that occur between the images. The proposed system uses Histogram of Gradients (HOG) [11] to detect the changes between the frames. HOG is an image feature descriptor as well as a Human Feature descriptor. HOG descriptor is implemented by dividing of image into small cells and the computation of the histogram of gradient directions is performed in each cell. Support Vector Machine (SVM) [12], classifier is well trained to recognize HOG descriptors. HOG is not influenced by noise, clutter, and illumination changes.

Algorithm 1: Occasionally moving object detection

Input: Video Frames F^t (A Sequence of Images) $t = 1, 2\ldots N$ ($N = $ Total Frames)

Output: Detection of Occasionally moving objects

Step 1: Generate Background Model with first few frames

Step 2: Get Background and use it as Reference Image (F^0)

Step 3: Extract HOG^0 in F^0 where HOG is a Histogram of Gradients

Step 4: For each frame $t = 1, 2\ldots N$

4.1: Extract HOG^t of all frames

4.2: Calculate $Corr^t i, j$ & $Diff^t i, j$ for current image F^t and reference image F^0 by using Equ. (2) and Equ. (3)

$$Corr^t_{i,j} = \frac{HOG^0_{i,j} \cdot HOG^t_{i,j}}{\left\|HOG^0_{i,j}\right\| \cdot \left\|HOG^t_{i,j}\right\|} \tag{2}$$

$$Diff^t_{i,j} = \left|F^t - F^0\right|, \tag{3}$$

Where Corr is Correlation and Diff is the difference

4.3: Extract the Area and Perimeter of Detected object to draw Bounding Box

4.4: Get Detection Binary Map $Binary^t_{(i,j)}$

$$Binary^t_{i,j} = \begin{cases} 1, \text{if } Corr^t_{i,j} < TCorr \&\& Diff^t_{i,j} > TDiff \\ 0, \text{Otherwise} \end{cases}, \tag{4}$$

Where $Binary^{t}_{(i,j)}$ is the Detection Binary Map value of (i, j) Pixel in frame F^t

Tcorr and TDiff are thresholds of Correlation and Difference maps

4.5: For each object detected in a frame, check the object in consecutive frames using (4.4)

4.6: If the detected object is static in consecutive long successive frames, then classify it as an occasionally moving object

Step 5: End For

3.4 Dynamic Uncommon Behavior Detection

The Proposed System detects the global abnormal human activities that occur across the frame by using Optical flow approach. Dense Optical flow algorithm [13] is used to calculate the Optical flow of frames to obtain the motion vectors. It is represented by (r, θ), where "r" represents the magnitude of each pixel and "θ" represents the deviation angle of each pixel from Eqs. (5) and (6)

$$r = \mathrm{sqrt}\left((x_2 - x_1)^2 + (y_2 - y_1)^2\right) \tag{5}$$

$$\theta = \tan^{-1}((y_2 - y_1)/(x_2 - x_1)) \tag{6}$$

In which (x_1, y_1) and (x_2, y_2) are initial and terminal coordinates of each pixel and r is the magnitude and (is the deviation angle of each pixel. Each frame is divided into A by B blocks in a uniform manner and find the Optical flow of all blocks in a successive manner.

Algorithm 2: Uncommon Human Activity Detection

Input: Video Frames f (A Sequence of Images)

Output: Detection of Uncommon Human activity

Step 1: For each frame f, do Preprocessing

Step 2: For each frame f, Estimate the motion values of each pixel value (i, j) by using Optical Flow

$$\text{Optical Flow} = (\mathrm{r}, \theta) \tag{7}$$

from Eqs. (5) and (6)

Where r- magnitude, θ-deviation angle of pixels

Step 3: Divide each and every frame f into A by B Blocks in a uniform manner

Step 4: Find Optical Flow for all blocks by using Eq. (8)

$$\mathrm{O_i} = 1/\mathrm{P} \sum \mathrm{Of}^{j}_{i} \tag{8}$$

Where O_i is the Optical Flow of i^{th} block, P is the Number of Pixels in a block, Of_i^j is the Optical Flow of j^{th} Pixel in i^{th} block

Step 5: Generate Motion Vector Map of all blocks in a successive manner

5.1: Assume M^i ($j \in K$) to Zero at the start of processing of each frame //K-Set of blocks

5.2: for all $i \in K$ do

5.2.1: $Th_d = \|O_i\| * BS$; // Th_d -threshold, O_i -Optical Flow of i^{th} block, BS-Block Size

5.2.2: $Fv_i/2 = O_i + \pi/2$;

5.2.3: $-Fv_i/2 = O_i +- \pi/2$; // Fv_i - Field view of Object i

5.3: for all $j \in K$ do

5.3.1: if $i \neq j$ then

5.3.1.1: Find the Euclidean Distance ED (i,j) between O_i and O_j

5.3.2: if $D(i,j) < Th_d$ then

5.3.2.1: Calculate the angle ϕ_{ij} between O_i and O_j

5.3.3: if $-Fv_i/2 < \phi_{ij} < Fv_i/2$ then

5.3.3.1: Assign $A = (-ED(i,j) / \|O_i\|)$

5.3.3.2: $M^j(O_i) = M^j(O_i) + \exp(A)$ (9)

5.3.4: end if for all

5.4: end for all

Step 6: Move TRAINING and TESTING Phase

3.5 Hybrid Clustering

The proposed framework uses hybrid combined clustering because it overcomes the disadvantages of hierarchical [14] and *k* means [15] approaches, which makes use of both features. Here the process starts by observing the movement of object vectors in blocks, and then each object is considered as initial cluster. The process goes on observing the features and spatial distance between the blocks to form clusters. It computes the vector matrix by matching the feature set and nearness in distances between the centroid of the clusters. The centroid value of the cluster is given in Eq. (10)

$$\text{Ck} = \sum_{i=1}^{n} x(k)/n \tag{10}$$

These are the advantages to construct the optimal number of clusters. By using the patterns of normal crowd activity, uncommon crowd activity is identified in testing phase.

Training and Testing Phase
In Training phase, the patterns of normal crowd behavior are identified by using hybrid clustering and they are saved as pattern blocks. Here each frame is divided into non colliding mega motion vector blocks N. Then compute Z_N by using Eq. (11)

$$Z_N = \sum H^j, \tag{11}$$

where Z_N represents the Motion Vector map value of Mega Blocks
H^j represents Motion Vector map value of blocks.

Perform the extraction of Space-Time features for each mega block to perform hybrid clustering. Perform hybrid clustering to get pattern blocks and set the centers as pattern blocks. Pattern blocks represent the behavior pattern of normal crowd activity.

In Testing, the Minimum Distance Matrix has to be calculated over all blocks is given below Eq. (12)

$$E(i, j) = \min\left|\left|f^{(i,j)} - w_k^{(i,j)}\right|\right|^2, \tag{12}$$

where E is the Minimum Distance Matrix, $f^{(i,j)}$ is the feature vector, $w_k^{(i,j)}$ is the Pattern blocks. Then highest value is observed among all the values, if the value is higher than a threshold ($\mathrm{Th_d}$) then classifies as abnormal frame.

4 Experimental Results

The Proposed system is developed using OpenCV-Python and the performance of the designed system is evaluated based on Pets [16], ABODA [17], UMN [18] and real time data sets. Pets and ABODA datasets are used for occasionally moving object detection and UMN dataset is used for uncommon human activity detection. The sample results are shown in Figs. 2 and 3 shows the occasionally moving object in indoor and outdoor scenario and abnormal human activity scenes. The performance is evaluated by considering the positive frames and negative frames taken from each dataset. The standard metrics of the classifier are calculated is shown in Tables 1 and 2 and resultant graph with existing system in Figs. 4 and 5.

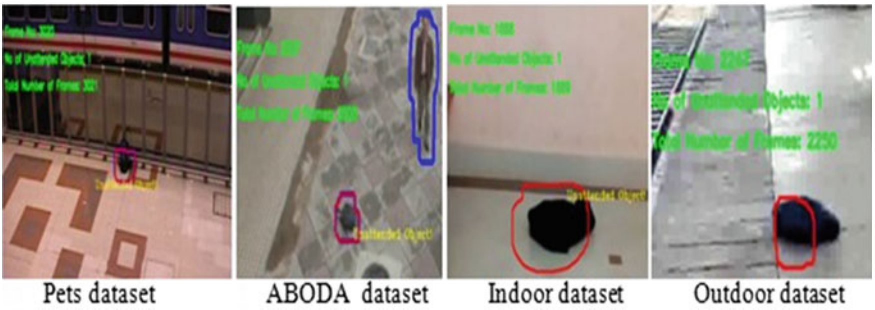


Fig. 2 The sample results for occasionally moving object with various datasets

Fig. 3 The sample results for uncommon human activity with various datasets

Table 1 The performance of the system—occasionally moving object detection

Algorithm	Sensitivity	Specificity	False alarm	Accuracy	Misclassify
Existing [19]	0.887	0.92	0.08	0.92	0.083
Proposed	0.89	0.932	0.068	0.95	0.05

Table 2 The performance of the system—uncommon human activity detection

Algorithm	Sensitivity	Specificity	False alarm	Accuracy	Misclassify
Existing [20]	0.904	0.907	0.093	0.933	0.066
Proposed	0.936	0.932	0.068	0.961	0.039

5 Conclusion

The detection of abnormality in the specific crowded areas is a difficult task. The main aim is designing the model to detect the abnormality in the human behavior either static or dynamic and to recognize the crowd event in the video sequence based on features extracted and the movement of motion vector pattern blocks. In this a new

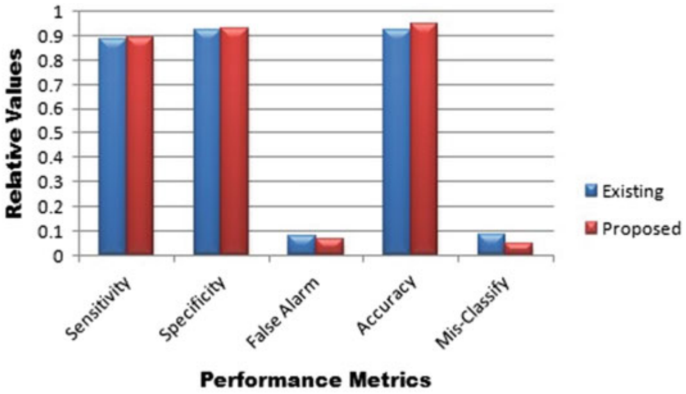


Fig. 4 Performance graphs of occasionally moving object detection

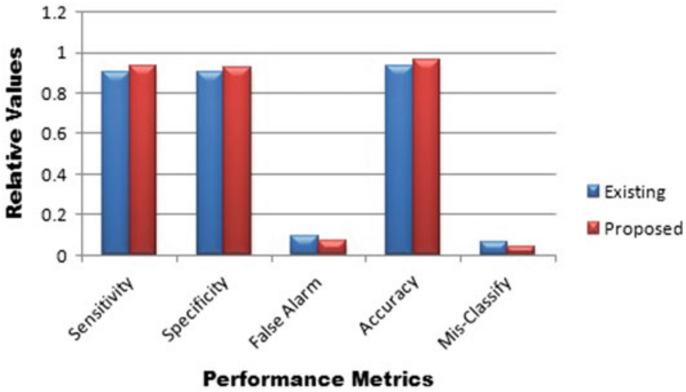


Fig. 5 Performance graphs of uncommon human activity detection

model is designed to identify peculiarity in crowd behavior involves in finding the displacement of motion vectors to form pattern blocks and HOG feature extraction and uses the hybrid clustering approach to construct the optimum number of clusters. It has been tested using standard PETS, ADOBA, and UMN dataset which showed accuracy of 96% in which processing time is lesser than the existing approaches. False alarm and misclassify is minimized. This work also tested with real time video shows it efficiency. In future the work is to extend various methods for different applications and to deal with the different events of multiple behaviors.

References

1. Lin K, Chen S-C, Chen C-S, Lin D-T, Hung Y-P (2015) Abandoned object detection via temporal consistency modeling and back-tracing verification for visual surveillance. IEEE Trans Inf Forensics Secur 10(7):1359–1370
2. Lin C-Y, Muchtar K, Yeh C-H (2014) Grabcut-based abandoned object detection. In: IEEE 16th international workshop on multimedia signal processing (MMSP), pp1–6
3. Grab cut segmentation to extract foreground, http://docs.opencv.org/3.1.0/d8/d83/tutorial_py_grabcut.html
4. Gaussian Mixture Learning Model, http://u.cs.biu.ac.il/~jkeshet/teaching/spr2017/lecture4.pdf
5. Jadhav LH, Momin BF (2016) Detection and identification of unattended/removed objects in video surveillance. In: IEEE international conference on recent trends in electronics, information & communication technology, pp 1770–1773
6. Jaouedı N, Boujnah N, Htıwıch O, Bouhlel MS (2016) Human action recognition to human behavior analysis. In: 7th international conference on sciences of electronics, technologies of information and telecommunications, pp 263–266
7. Kalman Filtering, https://en.wikipedia.org/wiki/Kalman_filter
8. Kumar S, John M (2016) Human activity recognition using optical flow based feature set. In: IEEE international carnahan conference on security technology (ICCST), pp 1–5
9. Optical flow, https://en.wikipedia.org/wiki/Optical_flow
10. Zhu S, Xia L (2015) Human action recognition based on fusion features extraction of adaptive background subtraction and optical flow model. Math Probl Eng, Article id:387464, 2015:1–11 (2015) http://dx.doi.org/10.1155/2015/387464
11. Histogram of oriented gradients feature extraction, http://mccormickml.com/2013/05/09/hog-peron-detector-tutorial/
12. Support Vector Machine (SVM), scikit-learn.org/stable/modules/svm.html
13. Gunnar Farneback Optical Flow, www.diva-portal.org/smash/get/diva2:273847/FULLTEXT01.pdf
14. Hierarchical Clustering, https://en.wikipedia.org/wiki/Hierarchical_Clustering
15. K-Means Clustering, https://en.wikipedia.org/wiki/K-Means_Clustering
16. Pets Dataset www.cvg.reading.ac.uk/PETS2006/data.html
17. ABODA Dataset imp.iis.sinica.edu.tw/ABODA/index.html
18. UMN Dataset http://mha.cs.umn.edu/
19. Lin Y, Tong Y, Cao Y, Zhou Y, Wang S (2017) Visual- attention-based background modelling for detecting infrequently moving objects. IEEE Trans Circuits Syst Video Technol 27(6):1208–1221
20. Lee D-G, Suk H-I, Park S-K, Lee S-W (2015) Motion influence map for unusual human activity detection and localization in crowded scenes. IEEE Trans Circuits Syst Video Technol 25(10):1612–1623

Scaled Conjugate Gradient Algorithm and SVM for Iris Liveness Detection

Manjusha N. Chavan, Prashant P. Patavardhan and Ashwini S. Shinde

Abstract With the extensive use of biometric systems at most of the places for authentication, there are security and privacy issues concerning with it. Most of the public places we usually visit are under reconnaissance and may keep track of face, fingerprints, and iris etc. Thus, biometric information is not secret anymore and sensitive to spoofing attacks. Therefore, not only biometric traits but also liveness detection must be deployed in an authentication mechanism which is a challenging task. Iris is the most accurate trait and increasingly in demand in applications like national security, duplicate-free voter registration list, and Aadhar program its detection must be made robust. Solution to the iris liveness detection by extracting distinctive textural features from genuine (live) and fake (print) patterns using statistical approaches GLCM and GLRLM are implemented. Popular supervised SVM algorithm and PatternNet neural network with second-order scaled conjugate gradient training algorithm are assessed. Both of these algorithms are found to be faster with PatternNet outperforms over SVM.

1 Introduction

With the advancement of technology and machine learning, biometric systems are adopted in various large-scale applications over typical password-based authentication. These systems based on physiological attributes of individual offer sophisticated and reliable authentication far better than systems which need the user to remember or carry some identity [1, 2]. Most accurate biometric trait reported among fingerprint, face, and Iris, is Iris which is increasingly in demand in applications like national security, duplicate-free voter registration list and Aadhar program in various countries as each individual's iris is unique, having complex texture pattern. In

M. N. Chavan (✉) · A. S. Shinde
E & TC Department, ADCET, Ashta, India
e-mail: manju3205@gmail.com

P. P. Patavardhan
E & C Department, GIT Belgaum, Belgaum, India

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_80

spite of these advantages, biometric systems over traditional methods are vulnerable to spoofing attacks which are either direct attacks at sensor level or tempering with database or storage media. Therefore, there is a continuous need for improvement and research going on to deal with spoofing attacks by distinguishing between the acquisition of genuine or fake images at the sensor level, inclusive of some additional response or intrinsic features of a person. Commercial iris technology based biometric authentication systems use algorithms developed by Daugman [3, 4] are not suffice and prone to spoofing attacks like printed iris images, replay attack or use of video, use of glasses or contact lenses. Therefore, there is a need for iris liveness detection to detect any kind of attack.

For liveness detection, many researchers developed new models/algorithms based on pupil dilation and or pupil dynamics, which though reliable require additional hardware, may take longer time for processing and also needs user cooperation [5–7]. Hardware-based techniques may not be suitable for detecting fake attack using unknown material and also if the user is wearing a cosmetic lens. The periodic pattern of dots is observed in case of spoofing attack with printed iris images which have reduced energy contents at higher frequencies [8]. Most of the time printed fake images used for spoofing, exhibits attributes like low quality, poor focus, blur, and poor contrast, unlike generic images which show quality features like sharpness, high contrast. Liveness detection will be carried out by making use of these attributes [9]. In recent years research using micro-textures is gaining attention as biometric traits are rich in textural information and can provide reasonably good results. Distinctive textural features form rich iris patterns like furrows, ridges and pigment spots are extracted to discriminate between live and fake samples [5]. Proposed work is a contribution in finding out a robust solution in analyzing textural pattern.

2 Related Work

Biometric systems used for authentication in security applications need to be reliable, simple and robust. To prevent and detect intruders or attackers in various biometric systems various methods are proposed so far. The multimodal approach provides good reliability and is resilient to specific attacks however it fails to detect vulnerable attacks, which are not considered at design time and may not work well for other unknown materials used for spoofing. Software-based spoofing methods are more flexible and reliable and can provide a generalized solution for spoofing attacks of any type by incorporating some intrinsic characteristics of biometric traits used. Biometric traits are characterized by textural information by extracting features using various descriptors. Many researchers have examined the discriminative power of descriptors on publicly available datasets using linear SVM classifier, which does not require parameter tuning and prior segmentation was not carried out K-means clustering with Euclidean distance for joint quantization of features is implemented and tested for percentage of fake/genuine misclassification. SID is found to be outperforming for both types of attacks like printed iris patterns and of the cosmetic contact lens. SIFT descriptors found to be computation intensive, SID, LCPD also shows moderate computation time simpler descriptors like LBP take much less time and suitable for low power applications like mobile phone authentication [1]. Deep learning techniques are nowadays seeking the attention of researchers, which have shown great success in computer vision tasks can be efficiently used to model the distribution of real/fake classes for liveness detection. Instead of relying on image quality features [9], deep learning framework learns representational and discriminative features. Though results are promising the approach is more computationally expensive than other existing solutions. Iris spoofing is examined for only specific printed and cosmetic contact lens dataset which is available publicly.

Iris textures are stochastic in nature with small repetitiveness. Variety of textures are available so a single operator is not adequate to describe. The operator should have qualities like efficient discrimination of different types of textures, robust to pose and scale variation and to spatial nonuniformity. Also, it should work well for small sample size and should have low computational complexity. The texture-based approaches reported to perform well for constrained databases with partial occlusion, uniform illumination, and high-quality input, therefore, there is need of feature extractor invariant to transformation and occlusions. Use of key point descriptors like SIFT and SURF improve the accuracy of iris recognition, it is found that SURF requires less time for identification compared to SIFT. It can be used for highly trust-based applications so useful for security [10]. For irises which textures are not distinct due to defocusing, blurring or occlusion by extracting local texture features and use of K-means clustering assessment time and hence system performance can be improved [11]. Using a spatial pyramid model and relational measure among neighboring regions, features are extracted in regions of different shapes which achieve improved performance for iris liveness detection for benchmark datasets [12].

3 Overview of Iris Recognition System and Feature Extraction

Typical iris recognition system as shown in Fig. 1 [3, 4]. Image acquisition is mostly carried out under constrained environment wherein eye image is captured by cameras under visible or near-infrared illumination. Captured irises are preprocessed by removing noise and reflections with the use of suitable filtering mostly median filtering and enhanced for further processing. Iris segmentation is the most important step, which is to approximately localize boundaries between iris-sclera and iris-pupil region. Eyelashes and eyelids are also localized and removed as they do not contribute any significant information in recognition. Poor segmentation would result in insufficient textural information which may lead to increased rejection rates and hence affects iris recognition accuracy. Normalization is carried out after segmentation to compensate for deformations like varying iris sizes, pupil dilation, and other factors. In the normalization process, iris rings are unwrapped into a rectangular

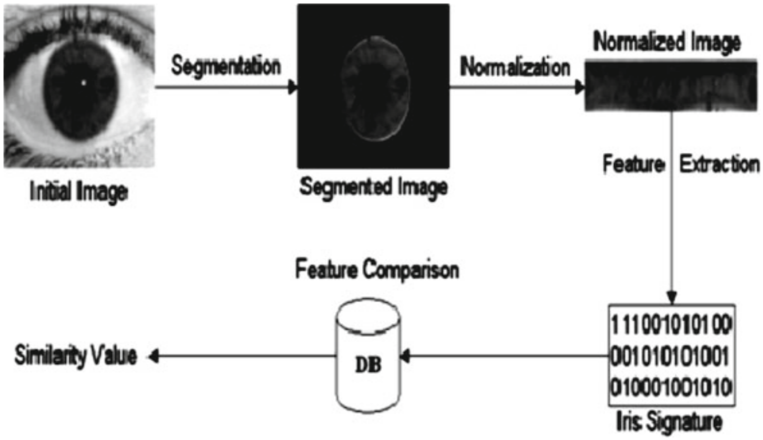


Fig. 1 Typical iris recognition system

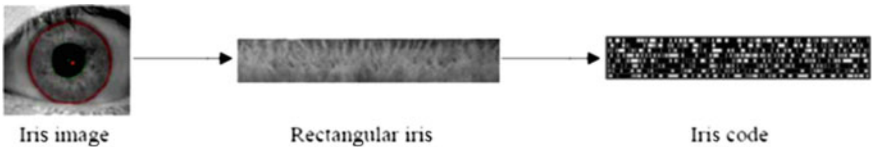


Fig. 2 Iris normalization using Daugman's rubber sheet model and generation of iris code from features

textural block of fixed size (20 × 240). Now after normalizing iris is ready for feature extraction [3, 4, 10] (Fig. 2).

Textural features of normalized iris are extracted which are unique for each individual. Extracted features are then encoded and iris code is generated for each registered iris and stored in the dataset. At verification and or matching stage enquired iris is processed through all these stages and code is compared against each iris code in the dataset and the user is either accepted as an authenticated user or rejected by the system.

In this paper, popular circular Hough transform is implemented as a segmentation algorithm and statistical textural methods gray-level co-occurrence matrix and gray-level run length matrix are used for feature extraction from normalized iris.

3.1 Proposed Work

Warsaw dataset consisting of eye images captured in NIR illumination. Fake images are created using laser printers with both low and high resolution on matt papers. Images which successfully spoofed the commercial iris recognition system are

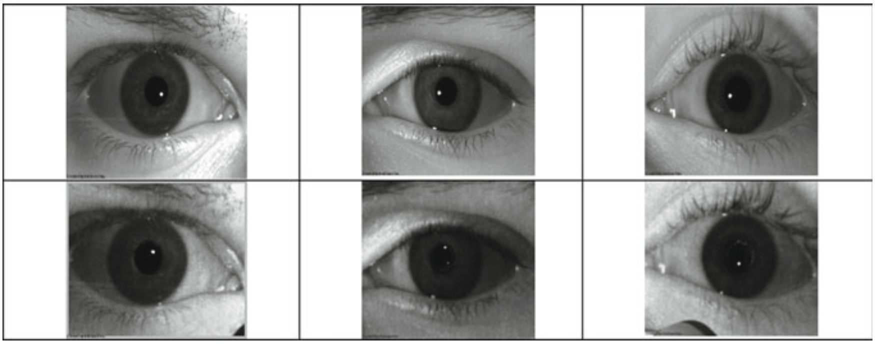


Fig. 3 Top row—Sample live iris images, bottom row—Sample print iris images

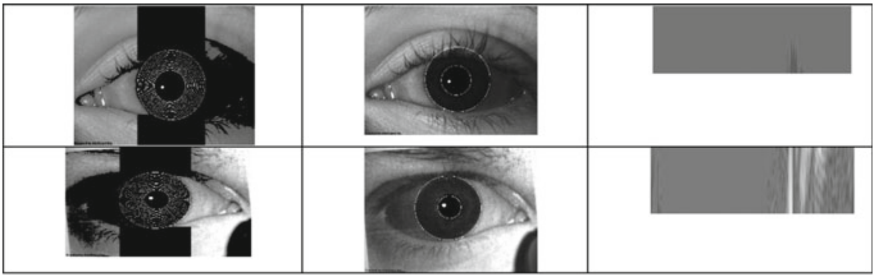


Fig. 4 Top row—Iris and pupil localization, segmentation and normalization for live iris Bottom row—Iris and pupil localization, segmentation, and normalization for print iris

preserved in the database. [12] Real and fake irises may reveal different optical characteristics under near-infrared lighting. These optical characteristics are helpful for iris liveness detection, which includes frequency distribution, Purkinje images, image quality features, and statistical texture features [13]. An optimization approach used in scaled conjugate gradient makes it effective over others as it does not contain any user-dependent parameters. It also avoids line search per learning iteration leading to the fastest algorithm among other gradient-based algorithms [14]. SVM is one of the most popular supervised machine learning algorithms used to analyze data and recognize patterns, used for classification with high performance and requires less tuning [5, 15, 16]. In view of these advantages of classification algorithms, SVM and scaled conjugate gradient (SCG) are chosen to test performance.

The proposed work is, therefore, to make use of statistical texture features like GLCM and GLRLM for liveness detection for Warsaw images. From normalized iris, 76 and 44 features are extracted using GLCM and GLRLM, respectively. 40 Real (Live) and 40 Print (Fake) iris images are tested using SVM and SCG algorithm for classifying between fake and genuine irises. Classification accuracy and confusion matrix for both classifiers are given in Results and discussion (Figs. 3 and 4).

3.1.1 Gray-Level Co-occurrence Matrix (GLCM)

Second-order texture information can be extracted using GLCM [5, 13]. It is a matrix whose rows and columns represent a number of distinct gray levels or pixel values in the image. It calculates how often a pixel with gray level (Intensity or tone) value *i* occurs either horizontally, vertically or diagonally to adjacent pixel *j* in the region of interest. GLCM represents the distance and angular spatial relationship over an image sub-region of a specific region of a specific size are transformed into the co-occurrence matrix space by a given kernel mask such as 3 3, 5 5, 7 7, and so forth. Normally, it is analyzed in four directions (e.g., Horizontal (0°), Vertical (90°), Diagonal (45° and 135°). It is denoted as pairs $P(i, j)$, P_0, P_{45}, P_{90}, and $P_{135,}$ respectively.

In 1973, Haralick et al. defined a set of 14 textural features which can be extracted from the co-occurrence matrix, and which contain information about image textural characteristics such as Angular Second Moment, Energy, Correlation, entropy, Inverse Difference Moment (IDM), Sum Average, Sum Variance, Sum Entropy, Entropy, Difference Variance, Difference Entropy, Difference average, information measure of correlation 1 and 2 [13, 14].

Most significant and relevant parameters used for textural features are energy, entropy, contrast, IDM, and DM defined below

Energy:

$$E = \sum_{i=0}^{G-1} [P(i)]^2 \tag{1}$$

Entropy:

$$H = -\sum_{i=0}^{G-1} P(i) \log_2 P(i) \tag{2}$$

Contrast:

$$\text{CTR} = \sum_{n=0}^{G-1} n^2 \left\{ \sum_{i=1}^{G} \sum_{j=1}^{G} P(i, j) \right\}; \quad |i - \text{j}| = n \tag{3}$$

Inverse Difference Moment:

$$\text{IDM} = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} \frac{1}{1 + (i - j)^2} P(i, j) \tag{4}$$

Directional Moment:

$$\mathrm{DM} = \sum_{i=1}^{G}\sum_{j=1}^{G} M(i,j)|i = j| \tag{5}$$

3.1.2 Gray-Level Run Length Matrix Features

Higher order statistical texture information can be computed using gray-level run length method. A set of consecutive pixels with the same gray level is called a gray level run. Gray level run length matrix describes the total number of occurrences of run length j at gray level i in a given direction θ. In order to extract texture features, GLRLM can be computed for any direction. Mostly five features are derived from the GLRLM introduced by Galloway. These features are: Short Runs Emphasis (SRE), Long Runs Emphasis (LRE), Gray-Level Non-Uniformity (GLNU), Run Length Non-Uniformity (RLNU), and Run Percentage (RP) [13]. Some of the features are described here.

Short Run Emphasis (SRE) is evaluated as

$$\mathrm{SRE} = \frac{1}{n_r}\sum_{i=1}^{M}\sum_{j=1}^{N}\frac{P(i,j)}{j^2} \tag{6}$$

where n_r is the total number of runs, M the number of gray levels, and N the maximum run length.

Long Run Emphasis (LRE):

$$\mathrm{LRE} = \frac{1}{n_r}\sum_{i=1}^{M}\sum_{j=1}^{N} P(i,j)\cdot j^2 \tag{7}$$

Gray-Level Non-Uniformity (GLN):

$$\mathrm{GLN} = \frac{1}{n_r}\sum_{i=1}^{M}\left(\sum_{j=1}^{N} P(i,j)\right)^2 \tag{8}$$

Run Length Non-Uniformity (RLN):

$$\mathrm{RLN} = \frac{1}{n_r}\sum_{j=1}^{N}\left(\sum_{i=1}^{M} P(i,j)\right)^2 \tag{9}$$

Run Percentage (RP):

$$\mathrm{RP} = \frac{n_r}{n_p}, \tag{10}$$

where n_p is the total number of pixels in the image.

3.2 Classification Using Scaled Conjugate Gradient (SCG) Algorithm and SVM

Feature set for all iris images under test is used to train Pattern net neural network with 50 neurons in hidden layer and SCG back propagation as training algorithm, which updates weight and bias values according to the SCG method. Basic backpropagation algorithm adjusts the weights in the steepest descent direction in which the performance is decreasing most rapidly along the negative gradient but does not necessarily produce the fastest convergence [14]. In the conjugate gradient algorithms, a search is performed along conjugate directions, which produces generally faster convergence than steepest descent directions. Maximum epoch used to train the network is 1000. Performance of the network for a given set of features GLCM, GLRLM, and combined feature set is tested with 10, 15, and 20 fold cross-validation. ROC curves and confusion matrix elaborates the overall performance of classification.

In SVM, the basic idea is to select hyperplane which is a line which best separates the input data points in space by their class as wide as possible. For 2 class problem, class 0 and class 1 can be completely separated by the line. The distance between the line and closest points is referred to as margin which is calculated as the perpendicular distance between the two [5]. These points are called support vectors and play an important role in defining the line and hence classifier. The hyperplane is learned from training data using an optimization procedure that maximizes the margin. In practice, real data cannot be perfectly separated with hyperplane and hence need some relaxation, which is called soft margin classifier. A tuning parameter C is introduced that defines the amount of violation of the margin allowed. C influences the number of support vectors used, smaller the value of C, the more sensitive the algorithm to the training data and vice versa. SVM algorithm is implemented using different kernels such as linear, polynomial, and radial. RBF kernel is found to be outperforming among all kernels. Performance of SVM classifier on feature set is tested with 10-, 15-, and 20-fold cross-validation. Overall classification accuracy is evaluated.

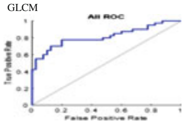
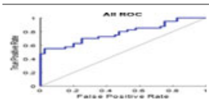
4 Results and Discussion

For 80 normalized iris images, 40 each class, 76 and 44 features are extracted in four different orientations using GLCM and GLRLM respectively. Combined GLCM–GLRLM features are a concatenation of the two individual features vectors. These features are used to train SCG and SVM classifiers with 10-, 15-, and 20-fold cross-validation. Experimental results given in the table describes confusion matrix of different K-fold cross-validation in terms of True positive, False positive, True Negative, and False Negative. From the table, it is evident that for GLCM overall individual class accuracy is better for 10-fold whereas for GLRLM it increases and maximum for 20-fold. The same is true for the combined performance of these two. SVM performs better for GLCM than the other two. For combined features, it performs worst. SCG shows poor performance for combined feature vector with tenfold. The overall accuracy of SCG is better than SVM.

5 Conclusion

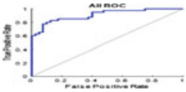
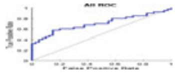
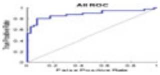
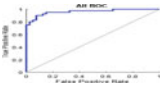
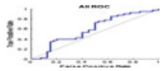
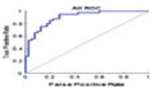
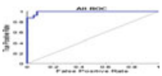
From experimentation, it can be concluded that overall accuracy for GLCM is better. SCG is able to outperform for iris liveness detection task than SVM. By increasing, data samples and features which distinguish between fake and genuine irises must be smartly chosen to improve performance. The overall accuracy of classification can be increased with accurate iris localization, preprocessing and feature selection. As some of the images in Warsaw database are of low resolution, it can be the cause of detection and hence classification performance which range from 60–90%. Future work will be in this direction also some feature reduction techniques will be employed in Table 1.

Table 1 Classification accuracy results of SCG and SVM

Total no. of iris images = 80, 40 live and 40 print images					Overall accuracy (%)	
SCG			SVM		SCG	SVM
GLCM  10 fold	$T_P = 35$	$F_P = 12$	$T_P = 29$	$F_P = 15$	78.8	72.5
	$F_N = 5$	$T_N = 28$	$F_N = 11$	$T_N = 25$		
	Sensitivity = 87. 5	Specificity = 70	Sensitivity = 72.5	Specificity = 62.5		
 15 fold	$T_P = 29$	$F_P = 12$	$T_P = 29$	$F_P = 12$	71.3	76.25
	$F_N = 11$	$T_N = 28$	$F_N = 11$	$T_N = 28$		
	Sensitivity = 72.5	Specificity = 70	Sensitivity = 72.5	Specificity = 70		

(continued)

Table 1 (continued)

Total no. of iris images = 80, 40 live and 40 print images					Overall accuracy (%)	
 20 fold	$T_P = 32$	$F_P = 14$	$T_P = 30$	$F_P = 15$	72.5	68.75
	$F_N = 8$	$T_N = 26$	$F_N = 10$	$T_N = 25$		
	Sensitivity = 80	Specificity = 65	Sensitivity = 75	Specificity = 62.5		
GLRLM  10 fold	$T_P = 30$	$F_P = 16$	$T_P = 14$	$F_P = 6$	67.5	63.75
	$F_N = 10$	$T_N = 24$	$F_N = 26$	$T_N = 34$		
	Sensitivity = 75	Specificity = 60	Sensitivity = 65	Specificity = 85		
 15 fold	$T_P = 35$	$F_P = 8$	$T_P = 16$	$F_P = 6$	83.8	61.25
	$F_N = 5$	$T_N = 32$	$F_N = 24$	$T_N = 34$		
	Sensitivity = 87.5	Specificity = 80	Sensitivity = 60	Specificity = 85		
 20 fold	$T_P = 36$	$F_P = 4$	$T_P = 18$	$F_P = 7$	90	60
	$F_N = 4$	$T_N = 36$	$F_N = 22$	$T_N = 33$		
	Sensitivity = 90	Specificity = 90	Sensitivity = 55	Specificity = 82.5		
GLCM+GLRLM  10 fold	$T_P = 9$	$F_P = 4$	$T_P = 29$	$F_P = 19$	56.3	65
	$F_N = 31$	$T_N = 36$	$F_N = 11$	$T_N = 21$		
	Sensitivity = 22.5	Specificity = 90	Sensitivity = 72.5	Specificity = 52.5		
 15 fold	$T_P = 33$	$F_P = 10$	$T_P = 32$	$F_P = 19$	78.8	58.75
	$F_N = 7$	$T_N = 30$	$F_N = 8$	$T_N = 21$		
	Sensitivity = 82.5	Specificity = 75	Sensitivity = 80	Specificity = 52.5		
 20 fold	$T_P = 37$	$F_P = 1$	$T_P = 31$	$F_P = 20$	95	66.25
	$F_N = 3$	$T_N = 39$	$F_N = 9$	$T_N = 20$		
	Sensitivity = 92.5	Specificity = 97.5	Sensitivity = 77.5	Specificity = 50		

References

1. Gragnaniello D, Poggi G, Verdoliva L (2015) An investigation of local descriptors for biometric spoofing detection. IEEE Trans Inf Forensics Secur 10(4):849–863
2. Unar JA, Seng WC, Abbasi A (2014) A review of biometric technology along with trends and prospects. Pattern Recognit 47(8):2673–2688
3. Daugman John (1993) High confidence visual recognition of persons by a test of statistical independence. IEEE Trans Pattern Anal Mach Intell 15(11):1148–1161
4. Daugman John (2004) How iris recognition works. IEEE Trans Circuits Syst Video Technol 14(1):21–30
5. He X, An S, Shi P Statistical texture analysis-based approach for fake iris detection using support vector machines. In: Proceedings of international conference on biometrics, pp 540–546
6. Lee E, Park K, Kim J (2005) Fake iris detection by using purkinje image. In: Advances in biometrics, ser. Lecture Notes in Computer science. Springer, vol 3832, pp 397–403
7. Kanematsu M, Takano H, Nakamura K (2007) Highly reliable liveness detection method for iris recognition. In: Annual conference SICE, pp 361–364
8. Jain AK, Bolle R, Pankanti S (1999) Personal identification in a networked society. Springer, New York, USA, pp 103–121
9. Galbally J, Ortiz-Lopez J, Fierrez J, Ortega-Garcia J (2012) Iris liveness detection based on quality related features. In: Proceedings of the 5th IAPR international conference on biometrics (ICB), pp 271–276
10. Mehrotra H et al (2012) Fast segmentation and adaptive SURF descriptor for iris recognition. Mathematical and computer modeling. Elsevier. https://doi.org/10.1016/j.mcm.2012.06.034
11. He Y et al (2012) Feature extraction of iris based on texture analysis. Advances in FCCS, vol 1, AISC 159. Springer, Berlin, Heidelberg, pp 541–546
12. Hu Y et al (2015) Iris liveness detection using regional features. Pattern Recogn Lett. https://doi.org/10.1016/j.patrec.2015.10.010
13. Sun Z, Tan T (2014) Iris Anti-spoofing. In: Marcel S et al (eds) Handbook of biometric anti-spoofing, advances in computer vision and pattern recognition. Springer, London. https://doi.org/10.1007/978-1-4471-6524-8_6
14. Haralick RM, Shanmugam K, Dijnstein I (1973) Textural features for image classification.IEEE Transactions On Systems, Man, And Cybernetics, November 1973
15. Huang X, Ti C, Hou Q-Z, Tokuta A, Yang R (2013) An experimental study of pupil constriction for liveness detection. In: Proceeding of IEEE workshop applications of computer vision (WACV), pp 252–258
16. He X, Lu Y, Shi P (2009) A new fake iris detection method. In: Proceeding 3rd international conference on advanced biometrics, pp 1132–1139
17. Pacut A, Czajka A (2006) Aliveness detection for iris biometrics. In: Proceedings of 40th annual IEEE international carnahan conference security technology, pp 122–129

Long Short-Term Memory-Based Recurrent Neural Network Approach for Intrusion Detection

Nishanth Rajkumar, Austen D'Souza, Sagaya Alex and G. Jaspher W. Kathrine

Abstract Intrusion detection is very essential in the field of information security. The cornerstone of an Intrusion Detection System (IDS) is to accurately identify different attacks in a network. In this paper, a deep learning system to detect intrusions is proposed. The existing recurrent neural network (RNN-IDS) based IDS is expanded to include Long Short term memory (LSTM) and the results are compared. The binary classification performance of the RNN-IDS is tested with various learning rates and using different number of hidden nodes. The results show that by integrating LSTM with RNN-IDS, the accuracy of intrusion prediction has improved against the benchmark dataset.

1 Introduction

Intrusion detection System (IDS) are hardware or software systems which can provide the ability to monitor networks for any type of unusual activity of intrusions such as misuse of access, unlawful access or login, hacking, etc. A very capable IDS must be able to identify and distinguish internal and external attacks. On detection of an intrusion an alarm is generated to alert the administrators or any software trigger can be used to block the further attacks. But even though intrusion systems of high detection have been deduced, the increased number of false positive attacks makes the development of any IDS technology unreliable.

N. Rajkumar · A. D'Souza · S. Alex · G. J. W. Kathrine (✉)
Karunya Institute of Technology and Sciences, Coimbatore, India
e-mail: kathrine@karunya.edu

N. Rajkumar
e-mail: rajkumar.nishanth@gmail.com

A. D'Souza
e-mail: austen.dsouza@gmail.com

S. Alex
e-mail: sagayaalex25@gmail.com

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_81

With the pervasive use of the Internet, more and more people are able to connect with each other through various means. This increase in connectivity means that there is more room for attackers to infiltrate a network that would have been secure a few years ago. The problem with identifying an attack is that we use known models to detect threats. However, the real difficulty lies in detecting an attack that is new, i.e., an attack that does not follow any well-defined pattern. This is where an IDS comes into play. An IDS can identify a security breach, that has occurred in the past, is currently ongoing or future attacks when they occur. This is done by classifying network traffic into the categories of "normal" and "anomalous". Network traffic is classified in two ways, i.e., the binary way as stated above or the five-category classification, i.e., Normal, Denial of Service (DoS), User to Root (U2R), Probe and Root to Local (R2L). The main objective of this work is to improve the correctness of the classification to effectively identify intrusive behaviour.

Techniques based on Machine Learning have been used to implement IDS but the inherent requirement of Machine Learning algorithms is that they need feature selection and belong to the shallow learning school. The notable downside of this is that these algorithms are unable to scale to the massive datasets that we would deal with in real-world scenarios. Moreover, they are not equipped for intelligent analysis and forecasting. Deep learning algorithms have the much better probability to generate models and extract much better impressions from the data.

This paper focuses on the following:

1. Implementation of RNN-IDS using LSTM. Focus is given on binary classification and the varying performance of the network based on the number of neurons and the learning rate.
2. The results are compared with RNN-IDs that does not use LSTM on the benchmark NSL-KDD dataset.

This paper aims to propose an effective intrusion detection system for networks and also to identify the framework to prevent the further intrusions in the networks. Section 2 contains the literature survey. Section 3 gives the framework for the proposed intrusion detection system.

2 Related Work

In current times, data mining and other data analysis based techniques are being used for analysing, identifying, and detecting intrusions. Nadiammai et al. [1] has proposed an IDS to detect the relevant and hidden data using data mining techniques which has led to less execution time. The proposed algorithm solves the major four issues such as data classification, need for decision-making by humans, short of data that is ready and labelled, and distributed denial of service (DoS) attack. Sabri et al. [2] have proposed an intrusion detection system which detects denial of service attack using techniques pertaining to data mining which includes clustering and classification. Dewa et al. [3] have proposed an intrusion detection system based on data mining which has achieved higher accuracy to new types of intrusion and a more robust performance is revealed when compared to traditional IDSs. Patel et al. [4] has proposed a paper where an amalgamated model is proposed to maximize the effectiveness in identifying the attacks that integrates the anomaly-based intrusion detection technique. Miller et al. [5] has done a research in which the IDS based on anomaly is used because it is more theoretically accurate in detecting the previously undocumented threats.

In recent years, deep learning is gaining more prevalence in identifying and mitigation intrusions in networks. Javaid et al. [6] have proposed a study in NIDS which helps to detect the network security breaches using deep learning approach. Chuanlong et al. [7] have proposed a paper for IDS based on recurrent neural network using deep learning approach. Shone et al. [8] has proposed a deep learning approach to intrusion detection in networks. Anomaly-based intrusion system assists in identifying both identified and unidentified intrusions. Jabez et al. [9] has proposed outlier detection for measuring anomaly in the dataset by using the neighbourhood outlier factor. Viegas et al. [10] has proposed a new evaluation scheme specific to the machine learning ID field. Umer et al. [11] have proposed taxonomy of flow-based IDS on the foundation of detecting maliciousness in flow record. Flow-based IDS is an inventive way of detecting intrusions in high speed networks.

Deep et al. [12] has proposed an intelligent system which will perform feature ranking on the origin of information gain and correlation. Dias et al. [13] has proposed a research paper for an IDS based on artificial neural network. Wang et al. [14] has proposed an effective IDS framework based on a support vector machine (SVM). Jha and Ragha [15] have proposed two ideas. First, they did the review in current statics of intrusion detection using support vector machine, and second, they proposed a novel approach to select the good feature for detecting intrusions. Abd-Eldayem [16] proposed an intrusion detection system which is based on Naïve Bayes classifier.

3 Proposed Methodology

On the basis of the literature survey it was decided to implement an IDS using Neural Networks. The neural network used was Recurrent Neural Network (RNN). The proposed framework for the IDS using Deep Learning is shown in Fig. 1. In addition to how a traditional RNN works, the model also implements Long Short-Term Memory (LSTM) to improve accuracy.

Long Short-Term Memory (LSTM) is the basic units required when constructing an RNN. Figure 1 shows the representation of LSTM. The most common construction of an LSTM cell consists of a cell, an input gate, an output gate and a forget gate. Each gate can be likened to an artificial neuron because they compute an activation of a weighted sum. The cell and the gates are connected. The cell is accountable for the “memory” of the LSTM. LSTM’s are short term memory that can last over a long period of time.

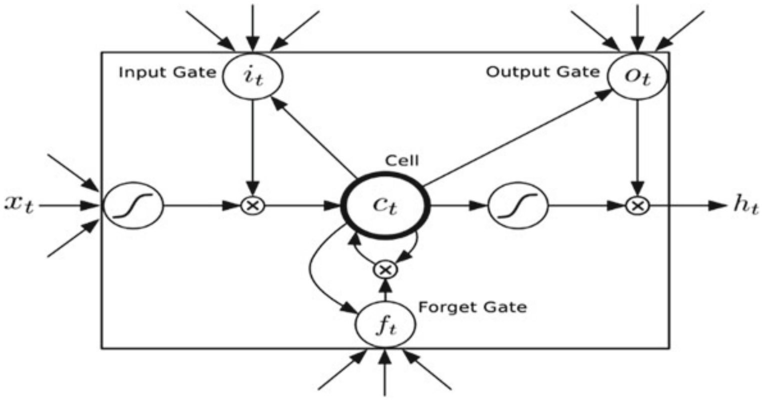


Fig. 1 Representation of LSTM

There is a debate existing over whether using LSTM improves the overall performance of a neural network. The consensus is that it works on a case by case basis. In this paper, an analysis is done to discover whether the use of LSTM is justified by augmenting its performance with the performance of the RNN. The Fig. 2 shows the architecture of the proposed and implemented Recurrent Neural Network (RNN) based IDS using Long Short-Term Memory (LSTM).

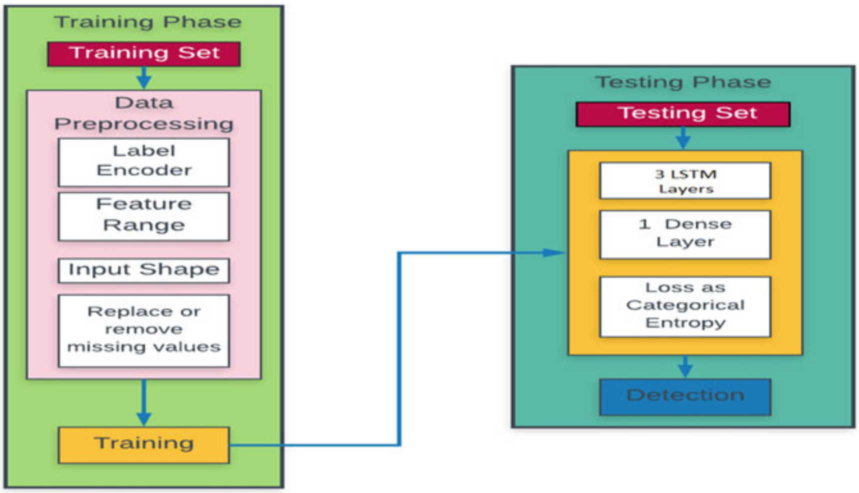


Fig. 2 Architecture of the proposed RNN-IDS using LSTM

A. Dataset

The NSL-KDD dataset was generated in the year 2009. This dataset is used in a wide range of IDS experiments and can be considered the benchmark dataset for these experiments. The NSL-KDD dataset solves the problem of the KD Cup (1999) dataset, wherein there existed many records that were inherently redundant.

Another advantage of the NSL-KDD dataset is that the number of records in the training and testing sets is better suited for the purpose of IDS testing. The KDDTest-21 dataset is a subset of the KDDTest+dataset and possesses a greater level of difficulty for classification. The dataset contains four types of attack records. The types of attacks are Dos (Denial of Service), U2R (User to Root), R2L (Root to Local) and Probe. Table 1 shows the composition of the dataset.

Every record (traffic record) is made up of 41 features and 1 class label. The features are further categorized as Basic (1–10), Content (11–22) and Traffic (23–41); numbered according to Table 1. Another unique feature of the dataset is that the testing set has some specific types of attacks that are not present in the training set. This acts like a more realistic basis for intrusion detection development and testing.

Table 1 Selected feature of the NSL-KDD dataset

No.	Features	Types	No.	Features	Types
1	Duration	Continuous	22	is_guest_login	Symbolic
2	protocol_type	Symbolic	23	Count	Continuous
3	Service	Symbolic	24	srv_count	Continuous
4	Flag	Symbolic	25	serror_rate	Conti nuous
5	src_bytes	Continuous	26	srv_serror_rate	Continuous
6	dst_bytes	Continuous	27	rerror_rate	Continuous
7	Land	Symbolic	28	srv_rerror_rate	Continuous
8	wrong_fragment	Continuous	29	same_srv_rate	Continuous
9	Urgent	Continuous	30	diff_srv_rate	Continuous
10	Hot	Continuous	31	srv_diff_host_rate	Continuous
11	num_failed_logins	Continuous	32	dst_host_count	Continuous
12	logged_in	Symbolic	33	dst_host_srv_count	Continuous
13	num_compromised	Continuous	34	dst_host_same_srv_rate	Continuous
14	root_shell	Continuous	35	dst_host_diff_srv_rate	Continuous
15	su_attempted	Continuous	36	dst_host_same_src_port_ra	Continuous
16	num_root	Continuous	37	dst_host_srv_diff_host_rat	Conti nuous
17	num_file_creations	Continuous	38	dst_host_serror_rate	Continuous
18	num_shells	Continuous	39	dst_host_srv_serror_rate	Conti nuous
19	num_access_files	Continuous	40	dst_host_rerror_rate	Continuous
20	num_outbound_cmds	Continuous	41	dst_host_srv_rerror_rate	Continuous
21	is_host_login	Symbolic			

B. **Data Preprocessing**

Label Encoding

Label Encoding is done to normalize the tags such that they contain only values between 0 and n − 1 classes. Label encoding is a method of normalizing the data to assist in easier classification. The label encoder also helps in numericalization of the data. Some features are non-numerical in nature and therefore need to be converted into numerical values based on the number of values they take.

Normalization

Certain features such as duration and src_bytes have a huge difference between the maximum and minimum values. So we use the Min Max scaler to normalize the dataset, where min and max are maximum and minimum values of each feature respectively.

The formula is

$$x_i = \frac{x_{i-\mathrm{Min}}}{\mathrm{Max} - \mathrm{Min}} \tag{1}$$

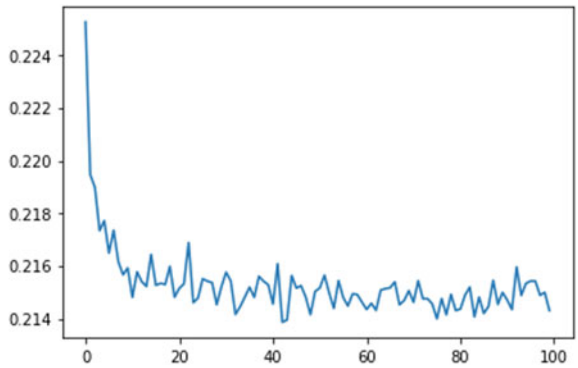
C. **Evaluation metrics**

An important indicator of performance is "acc" or accuracy of the lstm model on each dataset. The accuracy graph that is plotted is the accuracy of the model on each dataset for every epoch. The *x* axis is the epoch ranging from 1 to 100 and *y* axis is the accuracy ranging from 0.00 to 1.00. Another important indicator is the loss graph or "loss". Loss reviews the accuracy of a model by penalizing false classifications. The loss function that we used in this model was "mean squared error" or "MSE" since it is a classification model. MSE is a straight line between two points in Euclidean space. For classification models it is best to use mean squared error or cross entropy.

$$\text{MSE formula:} \quad \frac{1}{n}\sum_{t=1}^{n} \mathrm{e}_t^2 \tag{2}$$

When the LSTM model is trained on a dataset it returns a variable called history. The variable history contains various metrics stated while building the model. The history variable records the loss and accuracy for each epoch and graphs are plotted based on this. We can infer three different possibilities from the loss graph. They are underfit, overfit, and goodfit models. Underfit model is when it performs well on the train dataset and poorly in the test dataset. Overfit is when it performs well on training set whereas in test set it improves up to a point and then degrades. Goodfit is when the model performs well on the train and test set.

Fig. 3 Loss graph of the RNN-IDS on the dataset



4 Results and Discussion

In our research Keras deep learning framework based on TensorFlow is used. Binary classification is used and compared with the accuracy of LSTM Recurrent Neural Network with the fully connected Recurrent Neural Network model. The experiment was carried out on a Dell Inspiron 7559, which had the following configuration: Intel Core i7 CPU @ 2.6 GHz, 8 GB of memory and uses GPU acceleration. The GPU unit is an Nvidia GeForce 960 M GTX @ 1096 MHz.

The LSTM-IDS model we used has input shape of three dimensions (batch size, timesteps, features). In our model the input shape was (50, 10, 41). The entire dataset is reshaped to this 3D input since LSTM accepts only three-dimensional input. Since it is a binary classification we have two output nodes. The number of epochs used is 100. We used various combinations of hidden nodes and learning rate. No of hidden nodes were 20, 60, 80, 100, and 120. The learning rates used were 0.01, 0.1, and 0.5. The observations for the classification accuracy are shown in Table 2. In our experiment the RNN performs best at 60 hidden nodes and the learning rate at 0.1 as seen in Fig. 4. The experiments show that we obtain 74.33% for KDDTrain + dataset, 74.99% for KDDTest + dataset and 75.00% for Test-21 dataset and the time taken is 3258 s. Although the 100 hidden nodes have a higher accuracy at 0.1 learning rate, the accuracy is inconsistent and the loss graph is poor for this case. Hence 100 hidden nodes at 0.1 learning rate is not considered the best performance in our research experiment. Whereas the loss graph as shown in Fig. 3 for the case of 60 hidden nodes depicts that it is continuously decreasing showing that there is very little overfitting compared to 100 hidden nodes.

In Yin Chuan-long et al., the authors have shown the results obtained by fully connected RNN model without the use of GPU on binary and multiclass classification. The results gathered are compared with that of the binary classification. The model gives an accuracy of 68.55% for binary classification on the test-21 dataset. These results are based on the same NSL-KDD dataset. The performance of the LSTM-IDS is better in Test^{-21} dataset and the time taken is also much less in binary classification.

Table 2 Accuracy and time (s) of the RNN-IDS using LSTM

Number of hidden neurons	Learning rate	KDD train$^+$ (%)	KDD test$^+$ (%)	KDD test^{-21} (%)	Time (s)
20	0.01	65.66	65.15	64.25	3065
	0.1	67.51	63.19	63.55	3069
	0.5	34.94	34.89	34.78	3068
60	0.01	66.64	67.90	67.30	3253
	0.1	74.33	74.99	75.00	3258
	0.5	67.11	67.22	67.18	3280
80	0.01	66.81	67.31	66.88	3427
	0.1	51.01	51.03	50.90	3435
	0.5	34.93	34.83	34.84	3420
100	0.01	66.88	67.80	67.38	3530
	0.1	75.27	75.28	75.28	3541
	0.5	34.92	3482	34.82	3550
120	0.01	66.66	66.43	66.35	3857
	0.1	34.92	34.83	34.85	3852
	0.5	34.91	34.82	34.84	3865

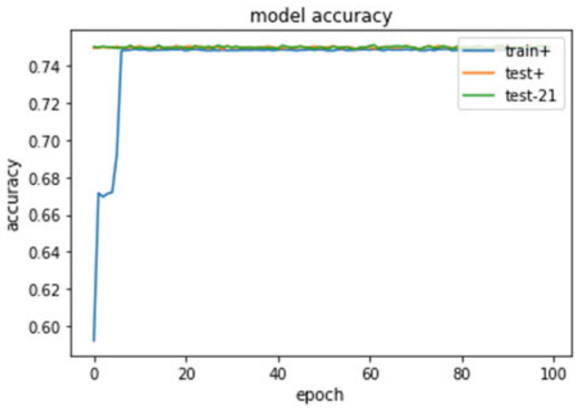


Fig. 4 Accuracy graph of the RNN-IDS on the dataset

In the observations from our research the results show that the RNN-IDS using LSTM model has a higher accuracy on the testing set than the training set as compared to RNN-IDS without LSTM. Moreover, it is seen to be consistently faster than a system that does not use LSTM. It also solves vanishing gradient problem that occurs in non LSTM models.

5 Conclusion

The RNN-IDS model using LSTM has a strong case to be used in intrusion detection, especially in binary classification. Compared to a non-LSTM-based RNN-IDS the results on the Test^{-21} set of the NSL-KDD dataset is very high. This model improves the ability to detect whether an intrusion is taking place or not. In future research, time optimization to better fit real time scenarios will be undertaken. Analysis to study the effect of different combinations of timesteps on LSTM model's accuracy and time taken have to be done to enhance the optimization.

References

1. Nadiammai GV, Hemalatha M (2014) Effective approach toward intrusion detection system using data mining techniques. Egypt Inf J 15(1):37–50
2. Sabri FNM, Norita NM, Seman K (2011) Identifying false alarm rates for intrusion detection system with data mining. Int J Comput Sci Netw Secur 11:4
3. Dewa Z, Maglaras LA (2016) Data mining and intrusion detection systems data mining and intrusion detection systems. (IJACSA) Int J Adv Comput Sci Appl 7(1)
4. Patel J, Panchal K (2015) Effective intrusion detection system using data mining technique. IJSRD—Int J Sci Res Dev 3(02)
5. Miller Z, Deitrick W, Wei H (2011) Anomalous network packet detection using data stream mining. J Inf Secur 2:158–168
6. Javaid A, Niyaz Q, Sun W, Alam M (2016) A deep learning approach for network intrusion detection system. In: Proceedings of the 9th EAI international conference on bio-inspired information and communications technologies (formerly BIONETICS), pp 21–26
7. Chuan-long Y, Yue-fei Z, Jin-long F, Xin-zheng H (2017) A deep learning approach for intrusion detection using recurrent neural networks. IEEE Access, https://doi.org/10.1109/access.2017.2762418, pp 21954–21961
8. Shone N, Ngoc TN, Phai VD, Shi Q (2017) A deep learning approach to network intrusion detection. IEEE Trans Emerg Top Comput Intell 2(1):41–50
9. Jabez J, Muthukumar B (2015) Intrusion detection system (IDS): anomaly detection using outlier detection approach. In: International conference on intelligent computing, communication & convergence (ICCC-2015), Procedia Computer Science, vol 48, pp 338–346
10. Viegas EK, Santin AO, Oliveira LS (2017) Toward a reliable anomaly-based intrusion detection in real-world environments. Comput Netw 127:200–216
11. Umer MF, Sher M, Bi Y (2017) Flow-based intrusion detection: techniques and challenges. Comput Secur 70:238–254
12. Akashdeep IM, Kumar N (2017) A feature reduced intrusion detection system using ANN classifier. Expert Syst Appl 88:249–257
13. Dias LP, Cerqueira JJF, Assis KDR, Almeida RC (2017) Using artificial neural network in intrusion detection systems to computer networks. In: Proceedings of computer science and electronic engineering (CEEC)
14. Wang H, Jie G, Wang S (2017) An effective intrusion detection framework based on SVM with feature augmentation. Knowl-Based Syst 136:130–139

15. Jha J, Ragha L (2013) Intrusion detection system using support vector machine. Int J Appl Inf Syst (IJAIS) 3:25–30
16. Abd-Eldayem MM (2014) A proposed HTTP service based IDS. Egypt Inf J 15(1):13–24

Video Frame Interpolation Using Deep Convolutional Neural Network

Varghese Mathai, Arun Baby, Akhila Sabu, Jeexson Jose and Bineeth Kuriakose

Abstract Video frame interpolation fuse several low-resolution (LR) frames into one high-resolution (HR) frame. The existing methods for video frame interpolation use optical flow method to determine motion in a scene, but computation using optical flow method is difficult, which can lead to artifacts in the output video. In many applications where we use video footages, there is a similarity in the content of footages. This similarity in content recommends that using some kind of context-aware approach can do better interpolation than the different existing interpolation techniques. We propose such a context-aware approach for video interpolation, the video frame interpolation using convolutional neural networks. In this proposed method, neighboring images are given as input to an end-to-end convolutional neural network which interpolates a frame between them. A comparative analysis of video interpolation technique using proposed RGB model and HSV model using metric standards such as SSIM, PSNR, and MSE is also included in the proposed method.

1 Introduction

Interpolation can be defined as a process of approximating the intermediate values of a continuous event from different discrete samples. In the real case scenario, video

V. Mathai · A. Baby · A. Sabu · J. Jose · B. Kuriakose (✉)
Department of Computer Science and Engineering, MITS, Ernakulam, India
e-mail: bineethbinz@gmail.com

V. Mathai
e-mail: varghesemathaime@gmail.com

A. Baby
e-mail: arunnellickamuriyil@gmail.com

A. Sabu
e-mail: akhilasabu1212@gmail.com

J. Jose
e-mail: jeexsonjose1996@gmail.com

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_82

interpolation is a technique which interpolates a frame between two neighboring frames. The dense motion between two consecutive input frames is estimated using stereo matching or optical flow algorithms and then interpolates different middle frames according to the dense correspondences estimation. This is the case happening in most of the existing frame interpolation methods. Any interpolation algorithm takes a pair of two images and produces an interpolated image. One of the main application of the interpolation algorithms is application which requires increasing frame rate [1, 2].

The most common and simple method for video frame interpolation is linear frame interpolation (LFI) [3], which simply takes a linear combination of neighboring frames to generate the target frame or the output. However, when different objects in the scene are moving over a stationary background, this technique fails. Motion compensated frame interpolation (MCFI) [4] is a popular, widely used and current industry standard algorithm for video frame interpolation, particularly in the applications concerned with upsampling of videos that is implemented in many modern HDTVs. But the problem with this techniques is that these methods are not context-aware. In many applications where we use video footages, the content of the video footages is similar. For example, consider CCTV camera footages, they have fixed background and only contains images of people moving across that background. Similarly, videos footages of football matches also mainly contain frames containing a ball and players on the green pitch. This similarity in content recommends that a context-aware approach could perform better at interpolation than the existing techniques. In this paper, we propose such a context-aware approach, video frame interpolation using deep convolutional neural network.

2 Methodology

2.1 System Description

In this method, input pairs of neighboring frames in a video are given to an end-to-end convolutional neural network which interpolates a frame between the input pairs. As a result, the frame rate of videos will be increased. For a given target frame, the corresponding inputs are the frames that are immediately before and immediately after the target frame. The input is a six-channel array composed of stacking the HSV data from the input video frames. We apply a convolutional layer on this data many times and then store the results obtained each time to ultimately arrive at the latent representation of the data. A set of filters is convolved with the image in a convolution step. In the deconvolution process, we pass the data through a deconvolutional layer, then concatenate the result with the similar sized layer from the convolution process before applying the next deconvolution. This allows the image regeneration process to have access to some details which may have been lost during the process of convolution [5]. The same process of interpolation is carried out by stacking the

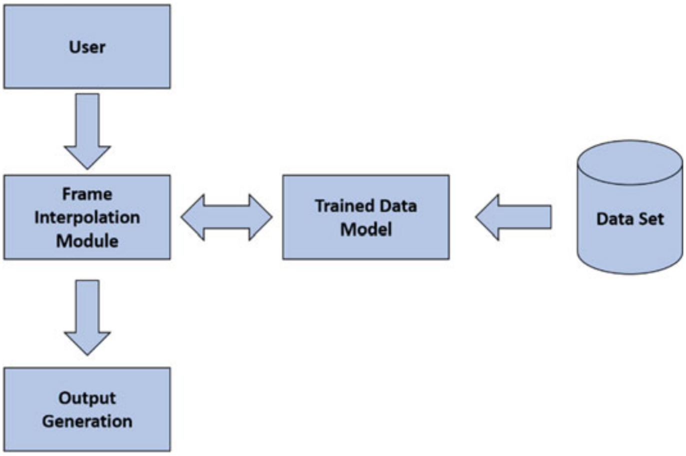


Fig. 1 Block diagram of the proposed system

RGB data from the input frames. Then, we will make a comparative analysis of both techniques using various metric standards.

Figure 1 gives the block diagram of the system which shows the working of the system. Initially, the user inputs the video to be interpolated. The neural network is trained to predict the intermediate frame based on the two input frames given. In frame interpolation, the input video is split into frames and is given as input to the trained network. Based on the previous inference, the intermediate frame is predicted. After generating frames for each set of input, all the frames are converted back to the video, and this is what happens in output generation. The RGB image data can be converted to HSV and can be used as input which increases the performance [6].

The system can be mainly divided into two parts, feed-forward neural networks and video frame interpolation module. The feed-forward neural networks is used to train the neural networks. We train the model on our training set for an epoch at a time. In frame interpolation, the first input video is converted to frames. From these frames, nth and $(n+2)$th frames are given as input to trained neural nets to generate intermediate frame. After generating all the frames, it is converted back to the video. The hardware used to enable the computation done in this study consists of a single NVIDIA 1080Ti GPU card.

Our python-based implementation of CNN for video frame interpolation enabled by the Keras and Tensorflow deep learning libraries is given in the following sections. The objective function we need to minimize is the Charbonnier error loss, given by

$$\rho(x) = \sqrt{\alpha^2 + \varepsilon^2}. \tag{1}$$

In this case, α represents the difference between the true frame f and the approximated frame f_i.

This function has various advantages over standard functions like the mean squared error loss as in this, the additional parameter ϵ lets us control the behavior of the function as the loss approaches zero. It is also possible to choose ϵ in such a way that when the loss falls below a certain threshold, the function increases quickly giving a pronounced signal that the model has converged within the acceptable given error range. For enabling gradient descent-based optimization, the Keras native Adam optimizer can be used with default parameter settings in all training cases, with a learning rate $\alpha = 0.0001$ for stochastic gradient descent (SGD) [7]. To avoid getting stuck at local optima, we can decrease the learning rate by a factor of 10 when the training loss becomes stagnant.

2.2 Data Collection and Processing

The dataset is created from a two and half hour duration movie. The entire movie is split into 15,000 frames using blender. The 9000 images are used as training dataset and 3000 frames are used as testing dataset. The videos are stored as individual frames with sequential numbering and PNG format. For a given video, there are grayscale as well as colored versions stored separately. First, a directory is chosen and all the files in that directory are enumerated. Then they are sequentially loaded into numpy arrays. We also set aside separate image sequences for validation in order to properly track the performance of our model. So, the remaining frames in the dataset can be used as validation sets.

2.3 Network Architecture

Ronnegerber et al. from Cornell University proposed U-Net in 2015 [8], which is a recursive CNN (RCNN) based network architecture that consists of a multi-layer contracting network. The network successively implements upsampling layers rather than pooling layers. This implementation increases the resolution and resultant feature space of the output data. The architecture is given in Fig. 2. The convolutional neural network architecture that we use consists of two main blocks. The first block type consists of two 3×3 convolutional layers, followed by a 2×2 max pooling layer. Similarly, the second block type contains a 2×2 upsampling layer, followed by two 3×3 convolutional layers. The entire network contains five blocks of the first type, followed by five blocks of the second type, in which the number of filters for each convolutional layer increases as the layers reach the center of the network, up to size 512. The input of the network has a shape of $6 \times 128 \times 384$, in which we concatenate the first frame f_i with the last frames f_j, on the channel axis. The final layer is again just a 1×1 convolutional layer with three filters, thus outputting a shape of $3 \times 128 \times 384$ because there is equal number of max pooling and upsampling layers.

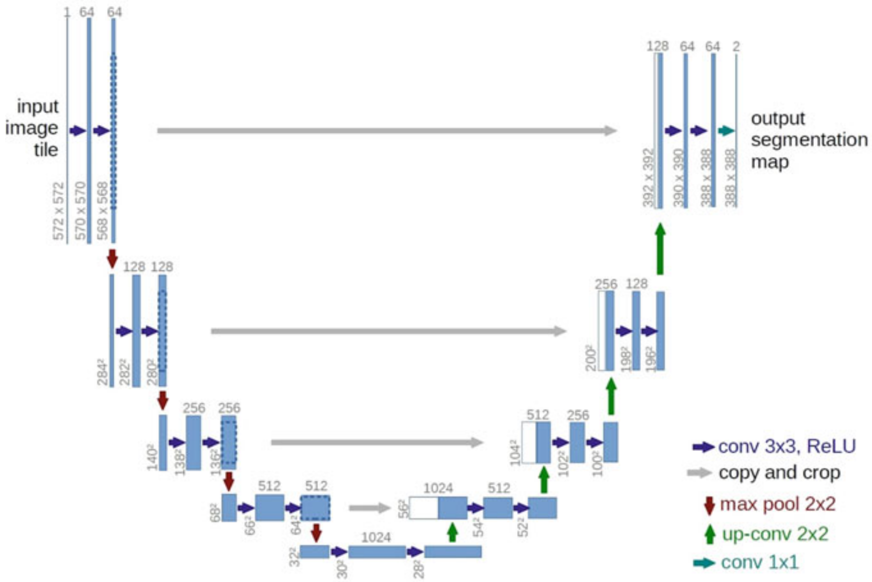


Fig. 2 UNet architecture used for the proposed system

This inherent symmetry of the network allows to increase performance. We have a merge layer before each block of the second type that concatenates the output of the opposite block in the first half of the network. This allows us to maintain feature information from the previous parts of the network, which had a thin field of view. As we use a 128×384 resolution image as our network input, and we have five max pooling layers, the height and width of the layer outputs reach a 4×12 resolution, that allows us to see a broader view of the input space as well. Along with all these, after each convolutional layer, we utilize the rectified linear unit (ReLU) nonlinearity activation function and to maintain the same input and output relative size, we also perform padding on the input in an orderly manner [9].

2.4 Network Training

The initial training can be done using the entire training set which takes over the course of some hours of time. The accuracy of the model can be increased by an increase in the loss incurred during validation, hence the weights defining the network topology are only updated when such an increase in validation loss occurs.

2.5 Metrics for Comparison

Along with analyzing the outputs from the system which implement video frame interpolation using CNN, the interpolated frames will be compared to their ground truths using various metrics such as Structural Similarity Index (SSIM), Peak Signal-to-Noise Ratio (PSNR), Mean Squared Error (MSE), and Mean Absolute Error (MAE) [10].

Mean Squared Error (MSE): The basic metric which we used to analyze the performance is the Mean Squared Error (MSE) between the generated frames and the ground truth images. Mean Squared Error gives the measure of the average of the squares of the errors or the deviations. It is basically the difference between the estimator and what is actually estimated. It is given by

$$\mathrm{MSE}(y, \hat{y}) = \frac{\|y - \hat{y}\|_2^2}{H * W * C} \tag{2}$$

where H, W, and C are the height, width, and depth of the input frames y and $\hat{y}$.

Peak Signal-to-Noise Ratio (PSNR): The PSNR is a log-scale ratio that is actually related to the Mean Square Error between an image and its ground truth. PSNR will penalize the noise that may get introduced to video frame during interpolation. It is given by

$$\begin{aligned}\mathrm{PSNR}(y, \hat{y}) &= 10 * \log_{10} \frac{\text{max power}^2}{\mathrm{MSE}(y, \hat{y})} \\ &= 10 * \log_{10} \frac{255^2}{\mathrm{MSE}(y, \hat{y})}\end{aligned} \tag{3}$$

where max power is the maximum possible pixel value of the image. When the pixels are represented using 8 bits per sample, this is 255.

Structural Similarity Index (SSIM): The structural similarity index (SSIM) is a metric that is used for predicting the perceived quality of images and videos. It is actually used for measuring the similarity between two images. SSIM is a qualitative metric and a good supplement to PSNR or MSE. Both the PSNR or MSE are robust quantitative measurements but is inconsistent with human visual perception.

$$\mathrm{SSIM}(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \tag{4}$$

3 Results

The input to the system is two frames of a video and the system is expected to produce a middle frame between the two input frames. The system generates the

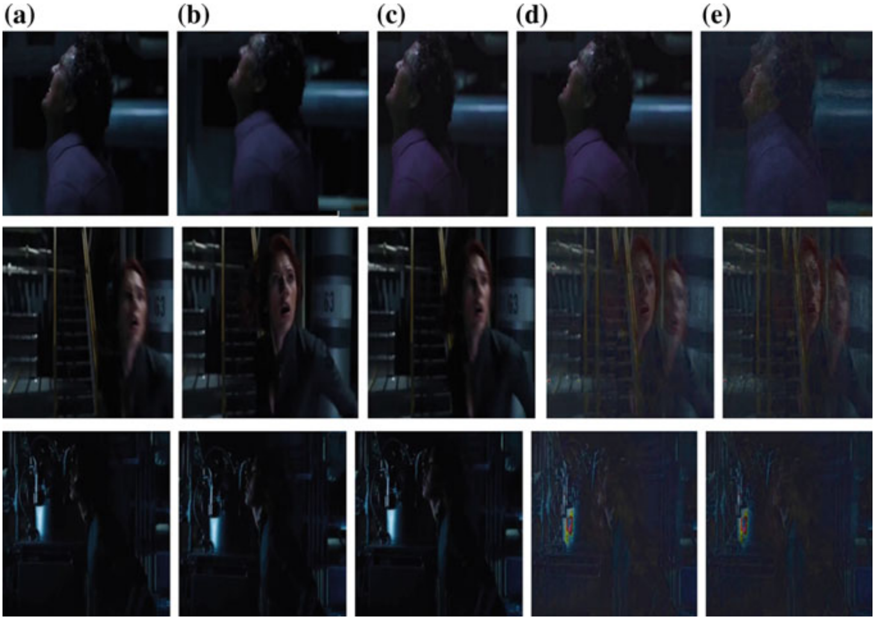


Fig. 3 Illustration of results obtained from three different movie scenes. **a** Frame 1. **b** Frame 3. **c** Ground truth. **d** Interpolated frame from RGB model. **e** Interpolated frame from HSV model

output based on the inference it made during the training phase. To estimate the effectiveness of the system, the video quality is used as the factor. Using the quality assessment metrics such as PSNR, SSIM, and the MSE, the video is compared to the ground truth. If the system does not produce output with expected accuracy, the architecture or training style has to be modified. Both RGB and HSV color models are used separately. In case of HSV model, the image was converted to HSV after it was read using OpenCV library using an inbuilt function. The HSV color space is more or less related to the way in which human beings perceive color whereas RGB color space defines the color in relation to the primary colors. The main reason behind the use of HSV version of video frame is because using the component Hue, we make the algorithm less sensitive to the variations of light. Also by using HSV instead of RGB, noise in the process output can be reduced. The result of our model will be less blurry on the moving object, thus we can expect a better SSIM. The results are shown in Fig. 3.

A comparative analysis of interpolation using RGB with the HSV is carried out using metric standards such as SSIM, PSNR, and MSE. The comparison of RGB model and HSV model using these metrics standards is given in Table 1.

The two models give the second frame by reproducing the structure successfully. But the noise is higher in the interpolated frame comparing with ground truth. The mean squared error is found to be lower for interpolation by HSV model compared to the interpolation by RGB models for all the set of frames we tested.

Table 1 Comparison of RGB model output with HSV model using quality metrics MSE, PSNR, and SSIM

Inputs	MSE		PSNR		SSIM	
	RGB	HSV	RGB	HSV	RGB	HSV
Frame set 1	436.94	343.49	21.73	22.77	0.49	0.52
Frame set 2	679.57	550.11	19.81	20.73	0.38	0.40
Frame set 3	577.81	472.21	20.51	21.39	0.37	0.41

Therefore, HSV gives a more similar interpolated frame with ground truth compared with RGB model. Noise is less in the HSV output compared to RGB output as PSNR value is higher for HSV model. Here, the results of all the testing sets will have higher value of SSIM for HSV model than the RGB model. So the HSV model will generate high-quality frame than RGB as HSV model reproduces structure better than RGB model.

4 Conclusion and Future Work

The video frame interpolation using convolutional neural network exploits a convolutional neural network that takes input pairs of neighboring images and tries to interpolate a frame between them. One of the applications for such an algorithm is in the data compression and also in upsampling process. In both cases, the generated images were compared to a ground truth using different metric standards of PSNR, SSIM, and MSE. Our output shows good results and it is clear that the network excels at predicting the interpolated location of objects in the image scenes, even if the end result is of lower quality than desired. Also, we conclude that interpolation using HSV model is slightly better than the RGB model.

This video interpolation technique can be used in HDTV's to increase frame rate of the videos. Currently, MFCI is used for frame interpolation, which has disadvantages like soap opera effect and motion blur. The proposed system uses DCNN for frame interpolation which can produce more strong and error-free results compared to all existing techniques. The proposed system can be also used in data compression and upsampling in the case where video has to be sent through a low bandwidth channel. Another advantage of the technique is that it can be also used to increase the frame rate of CCTV footages thereby making CCTV footages more reliable. It can be also used in animation domain to increase the frame rate of animation videos thereby increasing the smoothness of the animation videos. The effectiveness of the system can be further increased by using generative adversarial networks. For achieving better performance, different layer architectures can also be considered for different types of video types.

References

1. Kang K, Ouyang W, Li H, Wang X (2016) Object detection from video tubelets with convolutional neural networks. In: IEEE conference on computer vision and pattern recognition (CVPR), pp 817–825
2. Dong C, Loy CC, He K, Tang X (2016) Image super-resolution using deep convolutional networks. IEEE Trans Pattern Anal Mach Intell 38(2):295307
3. Hung K-W, Siu W-C (2011) Fast video interpolation/upsampling using linear motion model. In: 18th IEEE international conference on image processing
4. Guo D, Lu Z (2016) Motion-compensated Frame Interpolation with weighted motion estimation and hierarchical vector refinement. Neurocomputing 181
5. Ghutke RC, Naveen C, Satpute VR (2016) Temporal video frame interpolation using new cubic motion compensation technique. In: IEEE international conference on signal processing and communications (SPCOM)
6. Saravanan G, Yamuna G, Nandhini S (2016) Real time implementation of RGB to HSV/HSI/HSL and its reverse color space models. In: IEEE international conference on communication and signal processing, 6–8 April 2016
7. Kingma DP, Ba J (2014) Adam: a method for stochastic optimization. CoRR abs/1412.6980 available at http://arxiv.org/abs/1412.6980
8. Ronneberger O, Fischer P, Brox T (2015) U-Net: convolutional networks for biomedical image segmentation, CoRR abs/1505.04597, available at http://arxiv.org/abs/1505.04597
9. Song BC, Jeong SC, Choi Y (2011) Video super-resolution algorithm using bi-directional overlapped block motion compensation and on-the-fly dictionary training. IEEE Trans Circuits Syst Video Technol 21(3):274–285
10. Kotevski Z, Mitrevski P (2009) Experimental comparison of PSNR and SSIM metrics for video quality estimation. In: Davcev D, Gmez JM (eds) ICT innovations. Springer, Berlin, Heidelberg

Sentiment Analysis of Twitter Data Using Big Data Tools and Hadoop Ecosystem

Monica Malik, Sameena Naaz and Iffat Rehman Ansari

Abstract Sentiment analysis and opinion mining help in the analysis of people's views, opinions, attitudes, emotions and sentiments. In this twenty-first century, huge amount of opinionated data recorded in the digital form is available for analysis. The demand of sentiment analysis occupies the same space with the growth of social media such as Twitter, Facebook, Quora, blogs, microblogs, Instagram and other social networks. In this research work, the most popular microblogging site 'twitter' has been used for sentiment analysis. People's views, opinions, attitudes, emotions and sentiments on an outdoor game 'Lawn Tennis' have been used for the analysis. This is done by analysing people's positive, neutral and negative reviews posted on Twitter. Through this it has been analysed that how many people around the world really like this game and how popular this game is in different countries.

1 Introduction

Microblogging websites have progressed to the level of being a source of a variety of information. This can be accredited to the nature of microblogs, wherein people can present their opinions, upload real-time pictures, post comments on a discussion related to current issues or review sentiment for a product they must have used. Recently, some companies have started polling such microblogs, so that they can conclude a general sentiment for their product. A challenge faced by them is to build a technology that helps them identify and compile overall sentiment.

M. Malik · S. Naaz (✉)
Department of Computer Science and Engineering, School of Engineering Sciences and Technology, Jamia Hamdard, New Delhi 110062, India
e-mail: snaaz@jamiahamdard.ac.in

M. Malik
e-mail: monicamalik17@gmail.com

I. R. Ansari
University Women's Polytechnic, Aligarh Muslim University, Aligarh 202002, U.P., India
e-mail: iffat_rehman2002@yahoo.co.in

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_83

Sentiment analysis is the same as opinion mining. Twitter is one of the social networking sites which is used by million number of people and it receives tweets in millions every day. And this data can be used for the business purpose or industrial purpose. Big data is datasets which are complex and voluminous and are not dealt with traditional processing application software. In this work, analyses of Twitter data is done in the Hadoop environment, and for that Cloudera has been used which provides Apache Hadoop based software. The clusters of nodes are formed for distributed processing. Tweets are in the form of comments which are the opinions, emotions or sentiments of the people. This Twitter data is gathered by using Twitter API (Application Program Interface) [1]. Later, after analyses, the system gives the required output in the form of neutral, positive or negative tweets. Further, the analysed data is represented in the form of graphs and pie charts using Tableau software for visualization.

2 Related Work

Natural level processing at various levels of granularity is used in sentiment analysis. The classification task can be carried out either at the document level [2] or at the sentence level [3, 4] or at the phrase level [5, 6]. The data on microblogging platform like Twitter has views and opinions of people on various subjects. This data is ever increasing and of diverse nature which poses a lot of difficulty in its processing. A lot of work has been done on sentiment analysis of Twitter data some of which can be found in [7–9].

One of the methods used for extracting sentiment data is distant learning and has been used in [7]. Here, the tweets finishing off with positive emojis like ':)' ':-)' are taken as positive and emojis like ':(' ':/' as negative. In this work, the models have been built using support vector machines (SVMs), Naive Bayes and MaxEnt classifiers and it has been shown that SVM performs better than the other two classifiers. They have used unigram and bigram model along with parts-of-speech (POS) features in the feature space. This paper proves that the unigram model performs better than every other model. Bigrams and POS features do not help specifically.

A similar distant learning paradigm is used by authors in reference [9] to gather data. They use an alternate method for the classification task: subjective versus objective. Twitter accounts of major newspapers like 'New York Times', 'Washington Posts' and the likes are crawled to obtain the objective data. For the data which is subjective in nature, they gather the tweets finishing with emojis as was done in [7]. In contrast to the reports obtained in [7], they report that POS and bi-grams both give better results. Both these methodologies work on the *n*-gram model. But the data they use for preparing and testing is gathered via search engine queries and hence may show some level of biasness.

This biasness has been eliminated in [10] which considers manually annotated data. The features used here also present better results as compared to the unigram baseline. Additionally, an alternative technique for data representation is used which provides much better results as compared to the unigram models. In this work, three type models are experimented, those three being feature based, tree kernel and unigram-based model. Feature-based model includes some features from past literature, and some newly proposed ones are used. A new tree design has been created for the tree-based model, and the unigram model is used as baseline. A unigram only fetches 20% over the chance baseline for both the tasks and is a hard baseline. Only 100 features are used in this work and it is able to provide results at par with those obtained by the unigram model that uses over 10,000 features. Both feature based and unigram model are good but not better than tree-based model when it comes to the performance.

Models can be tested in combinations, the combinations work 4% more efficiently than unigram baseline for both classification tasks. It is evident from various studies that Twitter specific features such as hashtags, emoticons etc add only a slight value to the analyser [6].

Sentiment classification on Twitter data has also been carried out in [11]. In this work, the data has been collected from three websites: Twitter Sentiment, Twendz and TweetFeel which provide real-time sentiment detection. The data taken from these sources which are noisy and biased in nature have been used to train the model. After this, tuning and testing have been carried out on 1000 manually labelled tweets each. They suggest the utilization of syntax features of tweets like retweet, link, punctuations, hashtags and exclamation marks along with features like prior polarity of words and POS of words. The outcomes here show that the features that combine prior polarity of words with their parts of speech give the best results for the classifiers considered. There is not much effect of the tweet syntax feature.

Sentiment analysis by Aue and Gamon [12] is executed on feedback data from Global Support Services survey in 2004. They have analysed the role of linguistic features like POS. They perform broad feature analysis and feature selection and show that abstract linguistic analysis features add to the precision of a classifier.

Positive and negative score is assigned to each word, which is calculated separately. This starts with one seed for calculating positive scores and one for negative score, respectively. General algorithm for label propagation is used in [13].

3 Sentiment Analysis of Twitter Data for Tennis

In this era of information where enormous amount of data is available on hand for decision-making it is very important to understand the data properly and extract the features which are important for analysis. Big data is nothing but the datasets that are not only big but also very high in velocity and variety, which is difficult to tackle with the traditional tools and techniques. And as we all know the amount of data is increasing day by day so much and for that, we need to study and provide solutions for such problem to handle and extract knowledge out of these datasets.

Also, valuable insights are needed by the decision makers from such rapidly changing data, that ranges from customer interactions to day-to-day transactions and social network data. And for that, we use the application of advanced analytics techniques which is used to provide value to the decision makers using big data analytics.

Here, in this paper, we aim to analyse opportunities provided by the application of big data analytics in various decision domains and we also analyse some of the different analytics methods and tools which can be applied to big data. We will analyse the views of people on an outdoor game 'Lawn Tennis'. What are their views about this game, how do they feel about this game and how popular this game is around the world.

This paper focuses on analysing one such micro-blog 'twitter' and would create models in order to classify 'tweets' into positive, neutral and negative sentiments.

4 Proposed Work

In this paper, twitter data is used that is 'the tweets' posted by the users on Twitter. This data has an advantage over previously used datasets is that the tweets represent a true sample of the actual tweets in terms of language use and content as in they are not filtered or modified and are collected in a flooded fashion. Only 1% of data is provided by Twitter for analysis. Two dictionary tools are used in this work.

First one is the hand-annotated dictionary for emoticons. This dictionary helps to map emoticons to their polarity and on Wikipedia, there are almost 170 emoticons listed with their emotional state. For example, we can say :) is used for happy which means positive and :(is used for sad which means negative [2–4].

Second is an acronym dictionary, which is collected from the web. This acronym dictionary has English translations of over 5000 frequently used acronyms. For example, lol is used for laughing out loud, and bff is used for best friend forever.

5 Result Analysis

Figure 1 shows the geographical representation of our datasets visualizing the popularity of tennis across the world. Contrary to the popular belief that tennis is only popular in urban parts of the world, this map shows the reality, surpassing the financial or growth index factors of a country or area of regions, and the reach of this enthralling sport. This shows the diversity in the fan base of the sport across the continents. This map has pie charts hanging over different countries representing the popular sentiment of the country for the sport.

Figure 2 is a three factors based line chart which focusses on the varying sentiments across the ten most active countries when it comes to tennis popularity. Each country's slope represents the countrymen's state of mind regarding their method of putting up

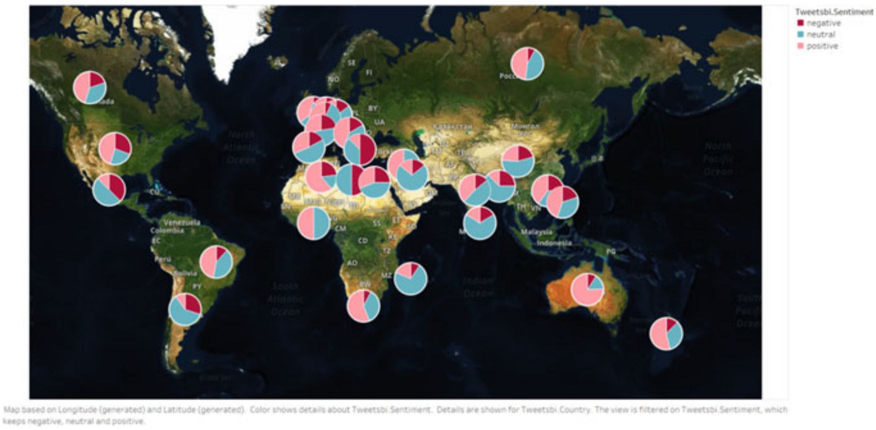


Fig. 1 Country-wise popularity of TENNIS

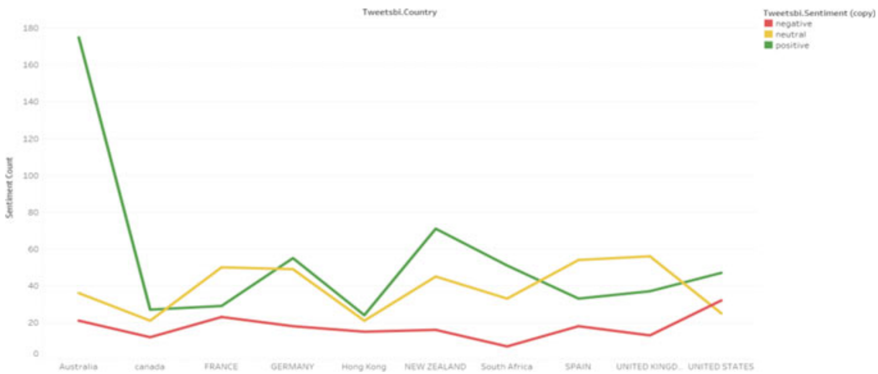


Fig. 2 Positive V/S Negative V/S Neutral slopes

opinions. For example—United States, very unbalanced nation, politically, shows the diversity in their opinions regarding the sport. There is no sheep walk, and everyone has a different set of opinions not influenced by anyone. Same goes to every other country except Australia, whose people are known for being chill and following their country's norms and traditions.

The pie chart in Fig. 3 shows the popularity of the sport across the world. From this analysis, we can draw two conclusions. First is, most of the people who use twitter as a social networking platform, use it to spread positivity and second is even if there is considerable amount of negativity around the tennis world in people's minds, they do not bother to show it to the world and they do not consider it worth their time and energy.

The ribbon chart depicted in Fig. 4 is useful for companies and industries dependent on tennis for their business. They can either shift their focus on to countries

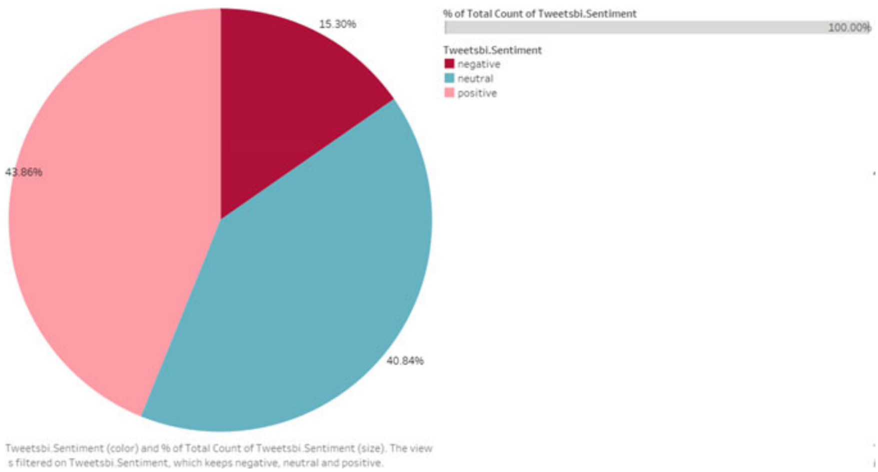


Fig. 3 Overview of the sentiments regarding ‘Tennis around the World’

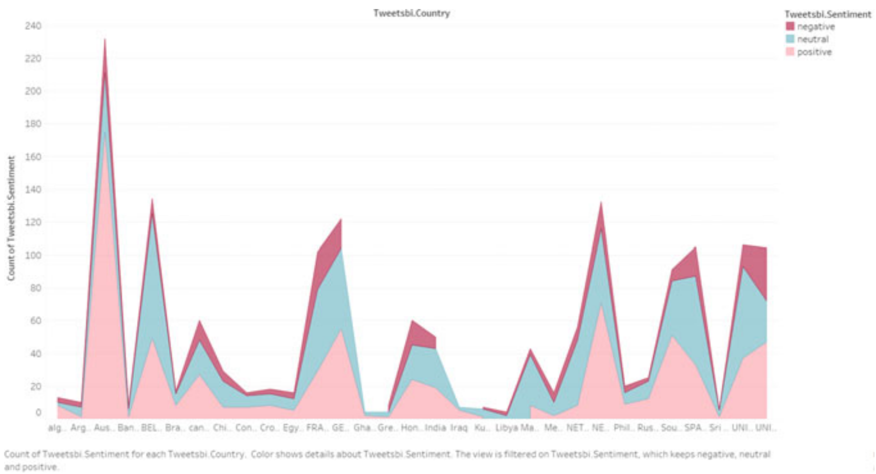


Fig. 4 Analysis of tennis popularity

that are not very developed but still have considerable amount of tennis popularity amongst the masses where tennis industry is still in infancy. Companies can also use this data to bank on to opportunities in rich nations where tennis popularity is already on its peak and the masses would accept any useful tennis related product with open hands.

6 Conclusion

Apache Hadoop is gaining a significant momentum from both industry and academia as the volume of data is growing rapidly. In this work Twitter data has been collected and analysed to see the sentiment of tennis among the people in the world. Hadoop has been used for this work due to its distributed architecture and ease of handling huge amount of data. It has been established from this work that contrary to the popular belief that tennis is only popular in urban parts of the world, it is equally popular in the rural areas. This analysis is very important for people who are involved in sports industry. The more the demand for this sport in a particular country the more their sale will be of the products related to the sport. So they can explore various parts of the world to see where their business can prosper most.

References

1. Ingle A, Kante A, Samak S, Kumari A (2015) Sentiment analysis of twitter data using hadoop. Int J Eng Res Gen Sci 3(6)
2. Pang B, Lee L (2008) Opinion mining and sentiment analysis. Now Publishers Inc., Foundations trends in information retrieval, available at http://portal.acm.org/citation.Cfm
3. Hu M, Liu B (2004) Mining and summarizing customer reviews. In: Proceedings of the tenth ACM SIGKDD international conference on knowledge discovery and data mining. ACM, pp 168–177
4. Kim SM, Hovy E (2004) Determining the sentiment of opinions. In: Proceedings of the 20th international conference on computational linguistics. Association for Computational Linguistics, p 1367
5. Wilson T, Hoffmann P, Somasundaran S, Kessler J, Wiebe J, Choi Y … Patwardhan S (2005) OpinionFinder: a system for subjectivity analysis. In:: Proceedings of hlt/emnlp on interactive demonstrations. Association for Computational Linguistics, pp 34–35
6. Bekkerman R, Gavish M (2011) High-precision phrase-based document classification on a modern scale. In: Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, pp 231–239
7. Go A, Bhayani R, Huang L (2009) Twitter sentiment classification using distant supervision. CS224 N Project Report, Stanford 1(12)
8. Bermingham A, Smeaton AF (2010) Classifying sentiment in microblogs: is brevity an advantage? In: Proceedings of the 19th ACM international conference on Information and knowledge management. ACM, pp 1833–1836
9. Pak A, Paroubek P (2010) Twitter as a corpus for sentiment analysis and opinion mining. In: LREc, vol 10, No. 2010
10. Agarwal A, Xie B, Vovsha I, Rambow O, Passonneau R (2011) Sentiment analysis of twitter data. In: Proceedings of the workshop on languages in social media. Association for Computational Linguistics, pp 30–38
11. Barbosa L, Feng J (2010) Robust sentiment detection on twitter from biased and noisy data. In: Proceedings of the 23rd international conference on computational linguistics: posters. Association for Computational Linguistics, pp 36–44
12. Aue A, Gamon M (2005) Customizing sentiment classifiers to new domains: a case study. In: Proceedings of recent advances in natural language processing (RANLP), vol 1, No. 3.1, pp 2–1
13. Zhu X, Ghahramani Z (2002) Learning from labeled and unlabeled data with label propagation

Detection of Chemically Ripened Fruits Based on Visual Features and Non-destructive Sensor Techniques

N. R. Meghana, R. Roopalakshmi, T. E. Nischitha and Prajwal Kumar

Abstract Nowadays great concern for everyone is health; hence primary requirement for sound health is eating good quality fruits. However, most of the available fruits in the market are ripened using hazardous chemicals such as calcium carbide, which is highly hazardous to human health. In the existing literature, less focus is given towards addressing the problem of identification of artificially as well as naturally ripened fruits, due to the complex nature of problem. In order to solve this problem, a new framework is proposed in this paper, which utilizes both the image features- and sensor-based techniques to identify whether the fruit is ripened by chemicals or not. By employing pH-sensor based techniques and visual features, it is possible to detect artificially ripened fruits and save the human beings from serious health hazards. The experiments were conducted and the results indicate that the proposed technique is performing better for the identification of artificially ripened banana fruits.

1 Introduction

The primary requirement for everyone is having good health condition, so eating good quality fruits provides sound health. The fruits are sweet-tasting plant product which contains fiber, water, vitamin C, and sugars. It also contains minerals, protein, cellulose, and various photo chemicals which protect human body against various disorders. Regular consumption of fruit is associated with anti-cancer, cardiovascular disease reduction, and declines aging factor. During the natural ripening process fruits attain desirable color, quality, flavor, palatable nature, and other textural changes during natural ripening process [1]. However it is quite impossible to get naturally ripened fruits, because most of available fruits in the market are ripened

N. R. Meghana (✉) · R. Roopalakshmi · T. E. Nischitha · P. Kumar
Alvas Institute of Engineering and Technology, Mangaluru, India
e-mail: meghanagowdanr@gmail.com

R. Roopalakshmi
e-mail: drroopalakshmir@gmail.com

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_84

using hazardous chemicals. For example, nearly 80% fruits like banana, mango, tomato, and papaya are nearly artificially ripened using different chemicals [2].

The process of fruit ripening is stimulated by applying artificial fruit ripening agents. Specifically farmers and vendors often use artificial ripening agents like calcium carbide and ethylene to control the rate of fruit ripening. Precisely, Calcium carbide is a corrosive and dangerous chemical which will be used as a ripening agent. Carbide consumption causes cancer due to phosphorous and arsenic poisoning which leads to diarrhea, burning sensation of chest and abdomen, vomiting, thirst, permanent damage of eyes, shortness of breath, weakness. Due to these reasons, as per PFA (Prevention of Food Adulteration) act in 1955, artificial ripening of fruits using calcium carbide is strictly banned.

Although calcium carbide is banned, farmers and traders are still using it for their own profit. Furthermore, due to this artificial ripening many agricultural products and fruits of Indian are rejected by few international market, which created a blackmark for all the Indian farmers as well as agricultural industry. Though India is the largest producer of bananas, its productivity has been reduced compare other countries like Guatemala, Indonesia. On the other hand, sensor-based non-destructive techniques are popularly used to detect calcium carbide in fruits. By employing sensor based technique it is possible to detect artificially ripened fruits and save the human from different hazards.

2 Related Work

Nelson et al. [1] presented a comparison of vitamin A, B, and C content of green, air, ethylene, and vine-ripened tomatoes respectively. Though this article indicates variations in the vitamin contents, but fails to detect the differences between the groups of air-ripened or ethylene-ripened tomato. Ahmad et al. [2] investigated the effect of temperature, ethylene, and their interaction towards the speed of ripening as well as quality of banana fruit. However it fails to differentiate between ethylene treated and non-ethylene treated bananas. F. J. Ramos et al. [3] estimated fruit firmness by using different strategies such as measurement of variables, acoustic responses, optical properties, analysis of impact forces, and nuclear magnetic resonance. Yet these techniques fail to predict whether fruit is artificially ripened or naturally ripened. K. de Mora et al. [4] presented an improved formulation allowing sensitive and accurate detection of less than 10 ppb arsenate in water.

In 2013, Zhao et al. [5] introduced a recognition system for artificial ripening, which extracted the image information of tomato using neural network based genetic algorithm, to identify the artificially ripened tomatoes. Even though, proposed technique achieved higher recognition rate, the complexity of the proposed algorithm is bit high. In 2013, Anmin Zhu et al. [6] proposed the concept of the neighborhood density to avoid the improper initial value and to initialize the cluster center which is based on the neighborhood data space correlation. In 2015, A. A. Bhosale et al. [7] introduced a capacitive sensing system, color index method and eco measurement to find various ripening stages of papaya. In 2016, Pratim Ray et al. [8] proposed an easy way to monitor the ripening stage of banana based on color indices and the information will be sent to the monitoring person automatically to the remote area. The information received will be processed by microcontroller to the authorize person using GSM module.

Very recently, in 2017, R. Karthika et al. [9] estimated an efficient image processing technique to detect the artificially ripened bananas. The author used threshold-based segmentation to extracted discriminatory features using Haar filter. In [10] proposed a method of Thermal Imaging Technique to detect whether fruit ripened by calcium carbide or not. This technique involves image pre-processing, segmentation, and feature extraction steps for processing of an image based on the infrared energy emitted by the fruit. Sahu et al. [11] developed an automated tool capable of identifying defects and maturity of mango fruits based on size, shape, and color features by digital image analysis. The author used MATLAB as the programming tool for classification and identification of fruits. The author failed to sort fruits based on mango quality which is essential for value addition.

To summarize, the existing systems fails to completely address the problems of identification of artificially and naturally ripened fruits. In order to overcome this proposed framework uses an image processing and pH sensor based techniques to identify whether the fruit is ripened artificially or naturally.

3 Proposed Framework

The sensor-based technique and image processing are combined to bifercate between naturally and artificially ripened bananas. The image processing involves feature extraction of banana image followed by finding the discriminatory behaviors for feature analysis. Figure 1 represents the block diagram of proposed method to classify between banana samples using image processing.

Proposed framework includes four modules: Image Enhancement, segmentation, feature extraction, classification. The image enhancement and segmentation of the image is made using K-means algorithm. Here the image is divided into discrete number of regions so that the pixels will have high contrast and similarity in each region. Followed by segmentation stage, feature extraction stage of the image is included which extracts statistical and texture features. More specifically features including variance, mean, contrast, standard deviation, energy, correlation, homogeneity, and gray-level co-occurrence (GLCM) matrix are extracted. In the classification module, SVM Classifier based on supervised learning, which is widely used for classification tasks is utilized. It consists of learning algorithms that analyze data and identify patterns, used for classification and analysis. SVM classifier distributes the data into different classes by the help of kernel function. The resultant extracted features are matched using SVM classifier which decides whether the fruit is artificially ripened or naturally ripened.

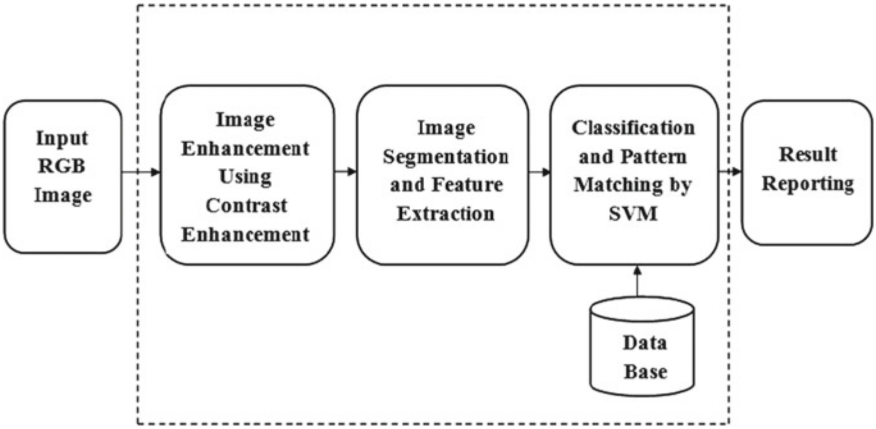


Fig. 1 Block diagram

As in case of pH sensor technique, the sample fruit is dipped into the warm water. The sensor is connected to aurdino board which in turn is connected to the PC. Based on the obtained pH value of that water it will decide whether the fruit is ripened chemically or naturally. After combining both image and sensor based technique obtained result will be displayed in the system.

Algorithm
K-means algorithm is used for the segmentation of input image. One of the unsupervised learning algorithms is K-means that solve the well-known clustering problem. The procedure is an easy and simple method to classify a given data set by a certain number of clusters. The main idea is to define k centers, one for each cluster.

4 Experimental Setup

To evaluate the proposed method pH sensor is used which is connected to the aurdino. For building electronics projects aurdino is used. It contains both physical programmable circuit board (often referred to as a microcontroller) and a software, or IDE (Integrated Development Environment) which is runed by the computer, used to write and upload computer code to the physical board. The aurdino IDE uses a simplified version of C++, making it easier to learn to program. pH is a measure of acidity of a solution the pH scale ranges from 0 to 14. pH sensor measure the level of pH in solution by measuring activity of the hydrogen ions in the solution. Below figure shows the snapshot of experimental setup of proposed framework (Fig. 2).

Experimental setup includes components like aurdino, ORP meter (pH sensor), Display, Camera. Here the sample fruit is dipped into the warm water. The sensor

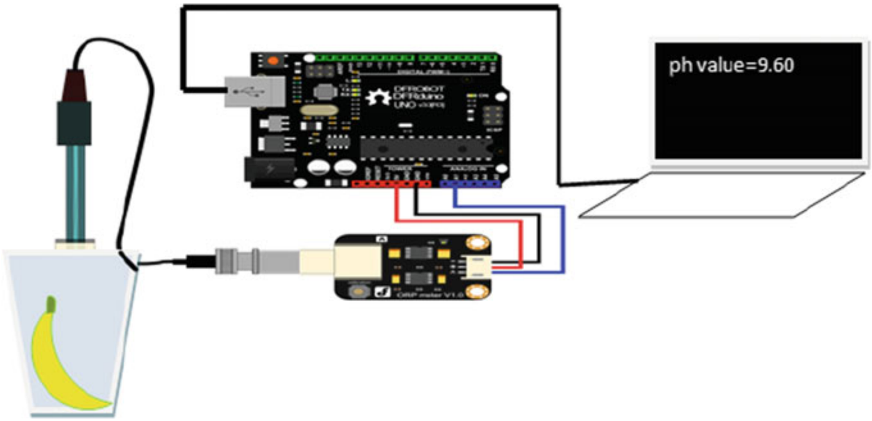


Fig. 2 Experimental setup of proposed framework

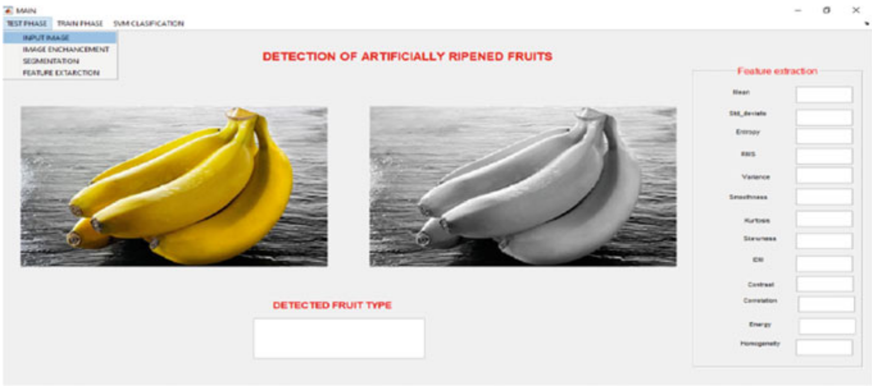


Fig. 3 Snapshot of the input image

is in contact with the sample and aurdino board which in turn is connected to the display.

5 Results and Discussion

Captured images are inputs to an image processing unit. Figure 3 shows the snapshot of input image. The important step in image processing is image segmentation. This step is to separate the image constituent areas of interest or regions. The processing and analysis of an image is based on segmented image. The segmentation based on color is used to segment the portion of banana fruit which is contaminated. Figure 4 shows the snapshot of the input image after segmentation.

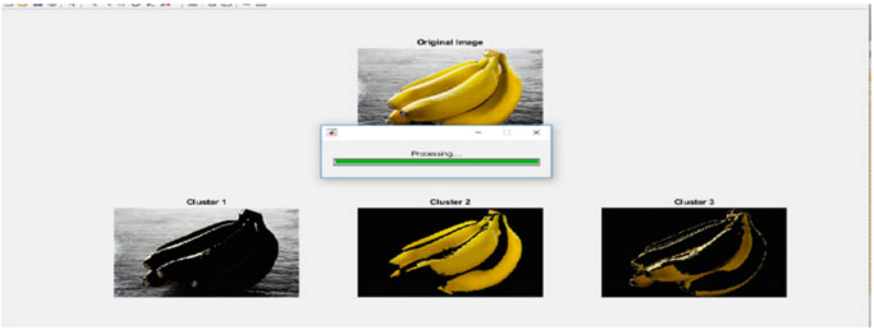


Fig. 4 Snapshot of the input image after segmentation

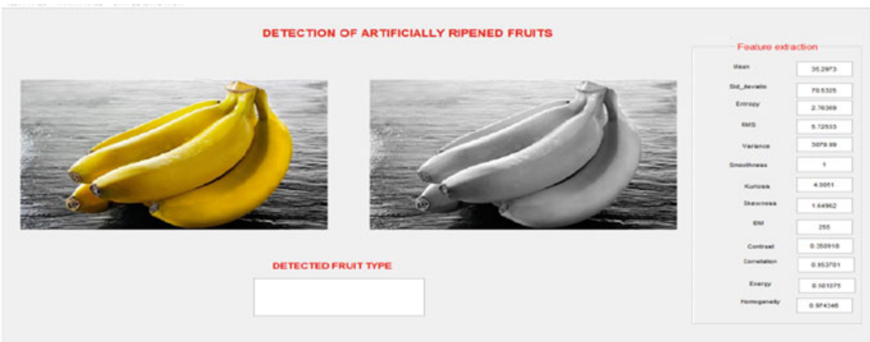


Fig. 5 Snapshot of the feature extraction

Image analysis is accomplished either straightly on the true image or either across individual color of plane. On the basis of intensity value of color of the image threshold is set because intensity values are sustained in homogeneous color region. Specific color range can be detected easily from whole image. The mean value of color is calculated and it is compared with each RGB pixels values of an image.

The next important step in image processing is feature extraction. The main aim of this is to extricate the feature from an image. It involves evaluating the area of an object in binary image. The input of an image may be logical or numeric. If it is numeric then nonzero pixels are appraised features like min and max, Area, Mean, Standard Deviation values of RGB are calculated for both the images and that will be stored in the database. Because all the extracted features have different values for different images. On the basis of Standard Deviation, Area, Mean, min and max values of RGB parameters of an image is compared with image stored in the database. Figure 5 shows the snapshot of feature extraction.

Figure 6 shows the snapshot of result obtained after image processing and pH sensor based technique. It concludes that the detected fruit is artificially ripened and its pH value is mismatched.

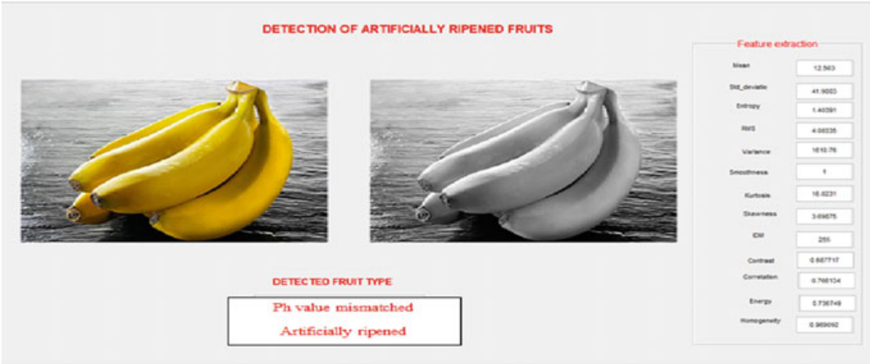


Fig. 6 Snapshot of the result

6 Conclusion

As the existing systems fail to completely address the problem of identification of artificially and naturally ripened fruits, the image processing and sensor-based techniques are proposed to identify the kind of fruit. The images which are captured will be the input to an image processing unit is used for identification of artificially and naturally ripened fruits. The processed image is compared with the database image to display whether the fruit is artificially ripened or not. By employing sensor-based technique it is possible to detect which kind of chemical component is present. So the conducted experiment indicates that the proposed method gives high accuracy compare to earlier techniques. In future, the proposed framework can be further improved for to chemically ripened fruits and value of the pH can be indicated.

References

1. Nelson M, Haber ES (1928) The Vitamin A, B, and C Content of artificially versus naturally ripened tomatoes. J Biol Chem 2:3
2. Ahmad S, Thompson AK (2001) Effect of temperature on the ripening behavior and quality of banana fruit. Int J Agric Biol 3:2
3. García-Ramos FJ, Valero C (2005) Non-destructive fruit firmness sensors. J Agric Res 3:61–73
4. de Mora K, Joshi N (2011) A pH-based biosensor for detection of arsenic in drinkingwater. In: Proceeding Springer, pp 1031–1039
5. Dhembare AJ (2013) Bitter truth about fruit with reference to artificial ripener. Proc Arch Appl Sci Res 5:45–54
6. Zhu A, Yang L (2013) An improved FCM algorithm for ripe fruit image segmentation. In: Proceeding of IEEE international conference on information and automation Yinchuan, pp 436–441
7. Bhosale AA, Sundaram KK (2015) Nondestructive method for ripening prediction of papaya. Proc Int Conf Interdisciplinarity Eng 19:623–630

8. Ray PP, Pradhan S, Sharma RK (2016) loT based fruit quality measurement system. In: Proceeding of IEEE international conference on green engineering and technologies, pp 224–229
9. Karthika R, Ragadevi KVM (2017) Detection of artificially ripened fruits using image processing. Int J Adv Sci Eng Res 2(1):20–34
10. Ansari Sheeba (2017) An overview on thermal image processing. Proc Second Int Conf Res Intell Comput Eng 10:117–120
11. Sahu D, Potdar RM (2017) Defect identification and maturity detection of mango fruits using image analysis. Am J Artif Intell 1(1):5–14

Dynamic Object Indexing Technique for Distortionless Video Synopsis

G. Thirumalaiah and S. Immanuel Alex Pandian

Abstract With a development of observation cameras, the measure of caught recordings extends. Physically dissecting and recovering reconnaissance video is work concentrated and costly. It is substantially more important to create a video description and the video can be observed in a good manner. So, here we describe a novel video outline way to deal with produce consolidated video, which utilizes a protest following technique for extracting imperative items. This strategy will create video objects and a crease cutting technique to gather the first video. Finally, output results that our proposed strategy can accomplish a high buildup rate while safeguarding all the imperative objects of intrigue. Hence, in this method, we can empower clients to see the synopsis video with high impact.

1 Introduction

In the decade, an expansive number of observation cameras have been sent and used as a piece of transportation focus focuses, ATMs, and various other open or private workplaces. Due to the reducing cost of sending cameras, it is fundamentally less requesting and more affordable to surveillance a specific zone. With the change of the web, a considerable number of surveillance chronicles are transmitted via the Internet. This would require seeing dynamically to select any key events and moreover perceive any suspicious conduct for a great deal of got video by associations and security affiliations. Regardless, with the help of high skilled persons, most perception accounts are being watched.

The advantages of the synopsis are used to test and validate the video objects and their timings in the scene. And the example in this domain is dynamic video method; it can extract important properties of the video. Another method is picture

G. Thirumalaiah (✉) · S. Immanuel Alex Pandian
Department of ECE, Karunya University, Coimbatore, Tamil Nadu, India
e-mail: tiru5502@gmail.com

S. Immanuel Alex Pandian
e-mail: immans@karunya.edu

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_85

retargeting [1, 2]; in this the selected frame is modified without changing the important movements can be preserved. This scheme of evaluation is implemented in the previously extensively. Including these methods some more schemes are utilized for video synopsis those are fast forward method [3, 4]. In that by skipping some content in the video, digested video is achieved.

So the condensed rate is very low and content loss is also possible. Another one is storyline method [5], in that important objects are extracted and present as storyline and content may be lost. Next to this content equidistant and time equidistant methods, in [6, 7] these methods, distortion is the major drawback. Hence, for solving these drawbacks and limitations, first extract the important parameters from the video background and present them in the space and take common background. Thusly, a solidified procedure video can be created. It should be seen that as opposed to various strategies, the combined video made by video once-over can express the aggregate stream of the scene. Additionally, video synopsis may change the relative arranging between items to lessen common abundance however much as could sensibly be normal. Some other authors have proposed different methods among that ribbon carving is another important aspect. In that, important aspects in the video are prompted and added to the single background scene, but the condensing ratio is very poor. In this paper, implementation of condensing ratio and content preservation is also possible.

2 Our Proposed Method

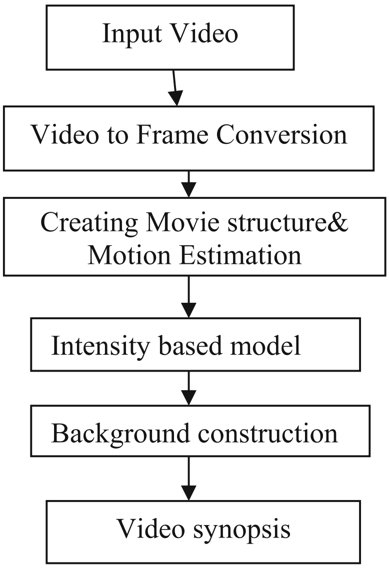
In conventional video summary methodologies, for the most part, static foundation pixels or casings are disposed of. Hence, the action of articles can be saved alongside the video being more minimal. The motivation behind video summary, which gives the watcher a reduced portrayal of the entire substance in the video, another video summation strategy. In this technique, the foundation pixels and the video objects are prepared. In this method, apllication of crease cutting to consolidate every movement. Exploiting crease cutting, key activities of the items are saved and excess developments are decreased. After these activities, a dense video can be produced. Figure 1 gives the process flow for our proposed method.

A. Object abstraction

In our object video summation technique, object abstraction is the primary step. As of late numerous protest following strategies have been presented [1]. These techniques can be straightforwardly connected in situations where a moving article may relate to an imperative occasion.

Moving shades are recognized and extracted by a simple method [8, 9] in which first the input video is converted into the frames and comparisons can be done between frames. Any value greater than 0, that current frame can be preserved and this procedure follows to last frame comparison. Figure 1 reflects complete process which is employed in this work for producing synoptic video. Figure 2 shows snapshots of tube getting and intensity flow of selected frames.

Fig. 1 Process flow



B. Object Condensing

In this segment, we use seam carving operator for extracted tube condensing. We can achieve this using spatial and temporal coherence cost functions.

(i) **Spatial Coherence cost**

In this, we are using seam carving operator which was proposed by shai avian et al. [10] by calculating the energy of the image we look forward to seams to be carved. We use forward energy idea of Rubinstein which improves the seam carving technique.

We start with the saliency map, which is used to identify the significant contents of the image. It is possible with the technique saliency region detection which is an easy and efficient one. The saliency regions can be found out by considering a range of frequencies of different pixels, i.e., using bandpass filters. The saliency map will high lighten the significant objects in the frame with perfect borders (Fig. 3).

In some cases, the edges were formed by the noise. This problem can be rectified by using canny detection. This technique uses gradients of sobel and finds the strong edges. The weak edges in the image are suppressed based on thresholding and the noise can be reduced by using Gaussian filter. The addition of saliency map and canny detected image gives the image with significant contents with strong edges so as to preserve them.

Seam carving algorithm is applied after calculating the compound energy. We are using dynamic programming for solving complex issues. This procedure requires more time and may not produce better results. So the frame undergoes quality checks before carving the seam right after seam removal computation. The similarity of

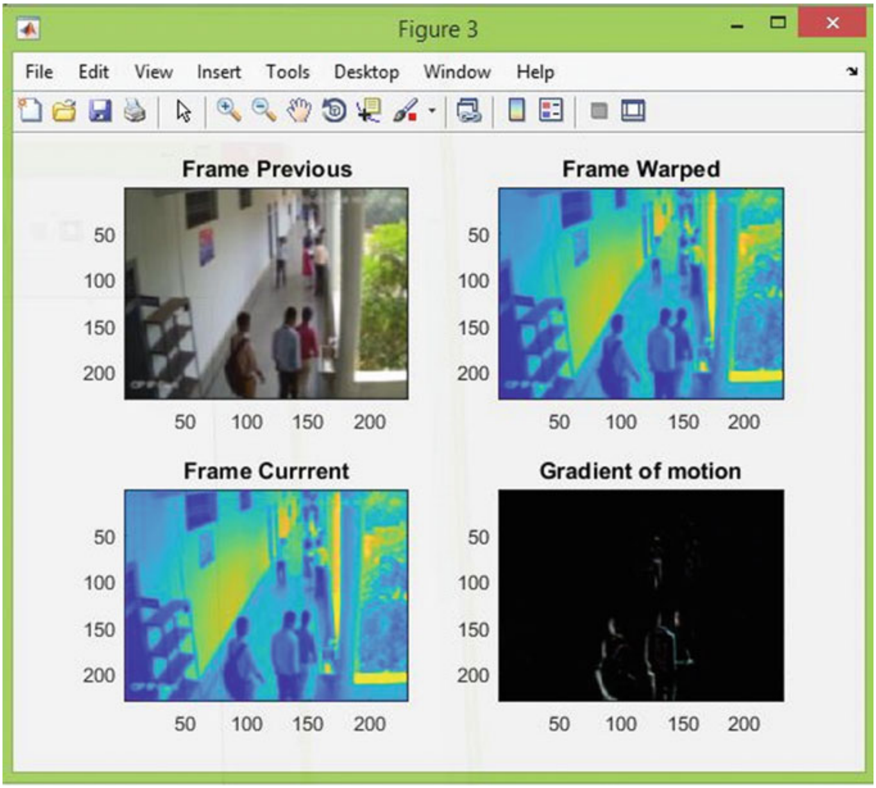


Fig. 2 Moving object detection (Testing frames from my own CCTV footage)

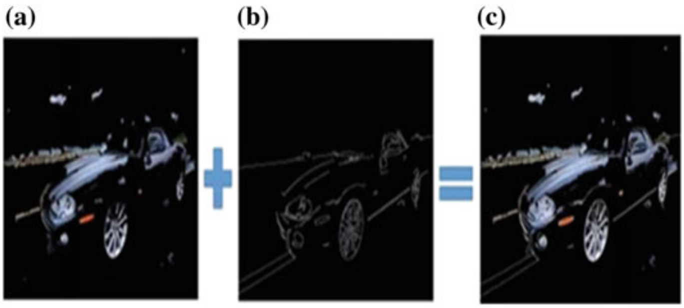


Fig. 3 **a** Saliency map **b** canny detection **c** area to be protected

processed frame and the original frame defines the cost. To measure the similarity we use bidirectional warping used specially in video retargeting. The cost may be increased due to the computation of similarity of the complete image. To reduce the cost, we calculate the window where the seam is removed and measure the similarity.

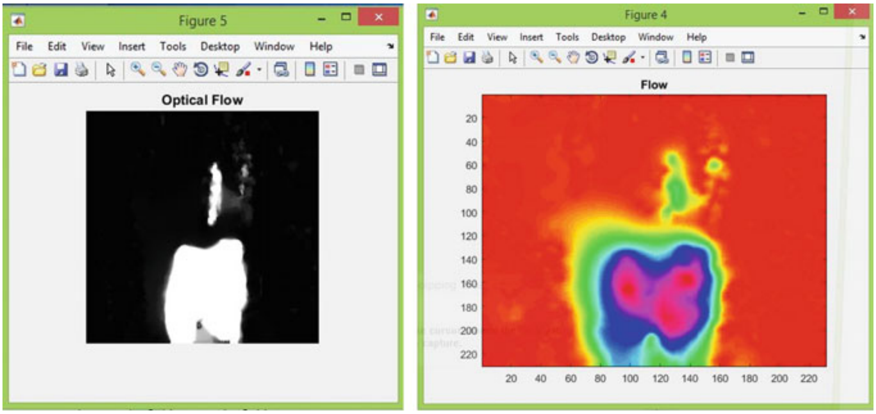


Fig. 4 Calculated seam and the cropping area for similarity check

There is no need to evaluate the similarity at the remaining positions because those are unaffected by the algorithm (Fig. 4).

(ii) **Temporal coherence cost**:

As said earlier the seam carving algorithm must be applied to individual frames because there may be chances of occurrence of artifacts if applied directly to a video. We identify the videos which are recorded from a static camera arrangement or dynamic camera arrangement.

In static camera arrangement, the recorded video background remains same throughout the video. For this, we calculated the sum of difference between the frames.

$$\sum_{i=0}^{l-1} F_i(x, y) - F_{i+1}(x, y) \tag{1}$$

The above Eq. (1) gives the difference map that consists of the overall motion of the object in the image and the seams calculated for the first frame are directly applied to next frames. Here the cost is effective.

(iii) **Energy Optimization**:

After carved video tracks generated, subtracted objects should be added to extracted background. Normally in surveillance video, the inner scene is in stable position. Before adding objects, the inner scene is generated using optical mean filter of the video.

The process of foreground adding to background to the condensed video is evaluated as shown in Fig. (3). The framework of this process can be evaluated using Eq. (2) (Fig. 5).

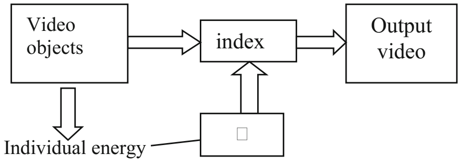


Fig. 5 Simple prototype model 66

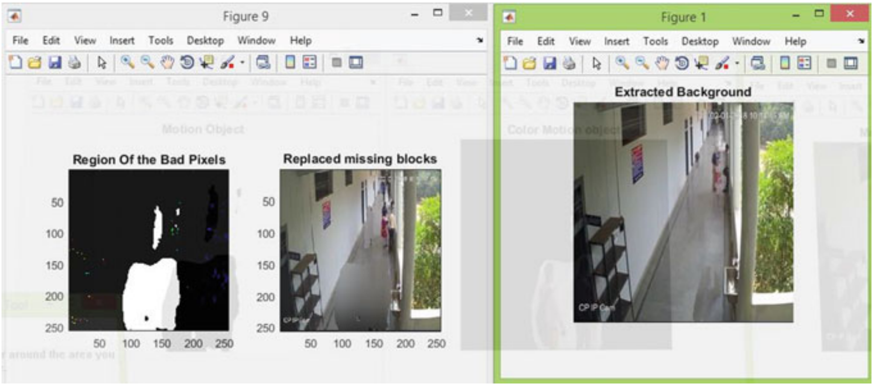


Fig. 6 Background construction of all extracted objects

$$E(M) = \alpha S_{\mathrm{as}} + \beta \sum\nolimits_{i} \mathrm{S}^{i}_{\mathrm{in}} + \gamma \sum\nolimits_{i,j} S^{i,j}_{\mathrm{pa}} \tag{2}$$

In the above equation, the energy of the objects is minimized by energy minimization method, where S_{as} is the object energy and S^{i}_{in} is the total frame average energy. Similarly the term $S^{i,j}_{\mathrm{pa}}$ is the total condensed frames energy. energy minimization is utilized to reduce the undesired energy objects and it can improve the efficiency of this method (Fig. 6).

By considering the huge computation of the more number of pixels and objects, optimization approaches are limited. Here, we use a greedy algorithm is used for better result. The position of each tube is defined as the term.

The evaluation of first frame number for every tube is assigned by this algorithm. Based on the assigned index, we can stitch each tube into the background and by doing this, we can generate the proper condensed video.

3 Experimental Results

To explain the accuracy and usability of the proposed scheme, five videos are used to evaluate the subjective performance: The Snooker game video, the office, Store video, and Bypass videos are used to explain the proposed method. The important

Table 1 Input video parameters

Name	Frame number	Resolution	Frame rate
Office.mp4	674	854 * 480	30 fs/s
Snooker.mp4	2400	400 * 224	20 fs/s
Store.mp4	1800	854 * 480	15 fs/s
Traffic.mp4	749	1920 * 1080	25 fs/s

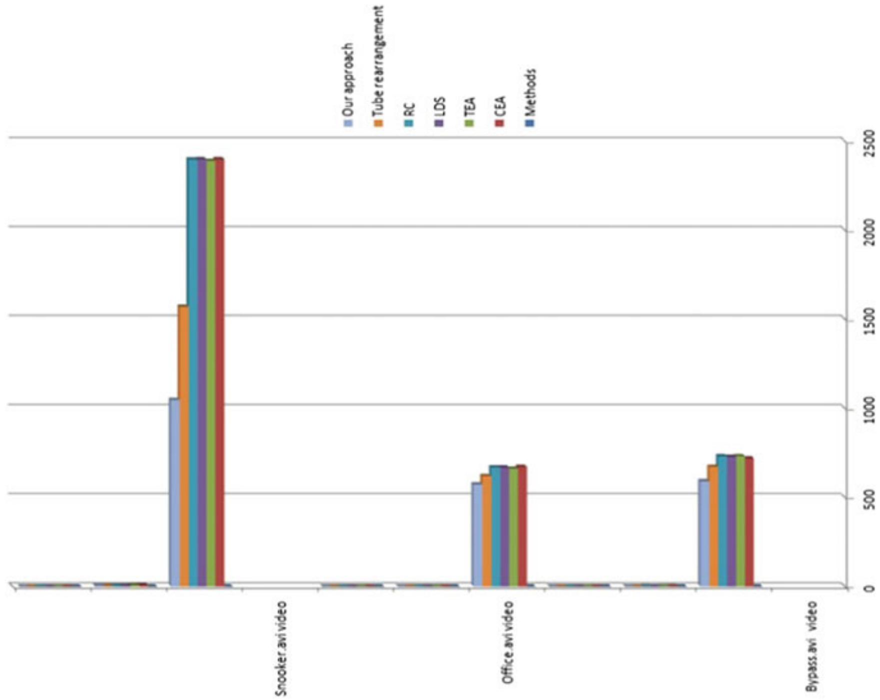


Fig. 7 Graphical view of comparative analysis

and useful parameters are listed in Table 1. All the programming codes are involved in this algorithm are related to either.mp4 or.avi formats. This is the most important and useful sole parameter of video analysis in our experimental condition in our paper. All the experiments are conducted on a laptop PC with Intel core i3, 2.5 GHz CPU, 8G RAM and under the WINDOWS 10 operating system.

As shown in Fig. 7, the performance of the moving objects tracking and extraction and video synopsis is achieved. The condensed and digested video itself is graphical representation of data which can be utilized to browse and extract the video content. So, the decrease in picture quality does not damage the summarized video.

Table 2 Comparison results for different methods

Methods	Corridor.avi video			Office.avi video			Snooker.avi video		
	Frames	Min distortion	Ratio	Frames	Min distortion	Ratio	Frames	Min distortion	Ratio
CEA [12]	720	1.5284	1.05;1	673	1.1625	7:05	2400	7.5501	4.2:1
TEA [12]	733	1.7035	2.12:1	663	1.2620	6:02	2391	7.3174	3.8:1
LDS	729	1.2888	5.13:1	669	1.0078	9:07	2400	5.0701	3.95:1
RC [13]	733	1.6258	1.14:1	670	1.1031	7:06	2399	5.0140	4.1:1
Tube rearrangement [13]	673	1.2735	2.34:1	621	1.1210	13.42:1	1572	4.7921	5.20:1
Our approach	592	1.2688	19.24:1	574	1.0732	14.84:1	1049	4.510	5.70:1

To assess the execution of the proposed approach, we have led a few analyses utilizing extraordinary observation recordings as unique video. We utilize three testing recordings from [11] for execution correlation. An evaluation and comparisons of these recordings are given in Table 1. With our proposed video summary approach, we extricate the moving objects from the video utilizing object following strategies and apply video crease cutting to the question cuts.

In this experiments results, three existing methods (content equidistant algorithm (CEA), time equidistant (TEA), noise model (LDS), and content rearrangement methods) [12, 13] are analyzed to evaluate our proposed method and its efficiency.

For producing video synopsis effectively, our proposed algorithm is employed on three different videos and parameters such as condensed frames, minimum distortion and aspect ratio are calculated and compared with the existing methods [12, 13, 2, 14]. These comparisons are shown in Table 2 and the snapshots of the tested videos are described with the help of Fig. 8.

In the last picture, a dense edge created by our technique is given, which can incorporate every one of these items. In this manner, our proposed technique can well protect every one of the items' action with a significantly more modest number of casings in the dense video.

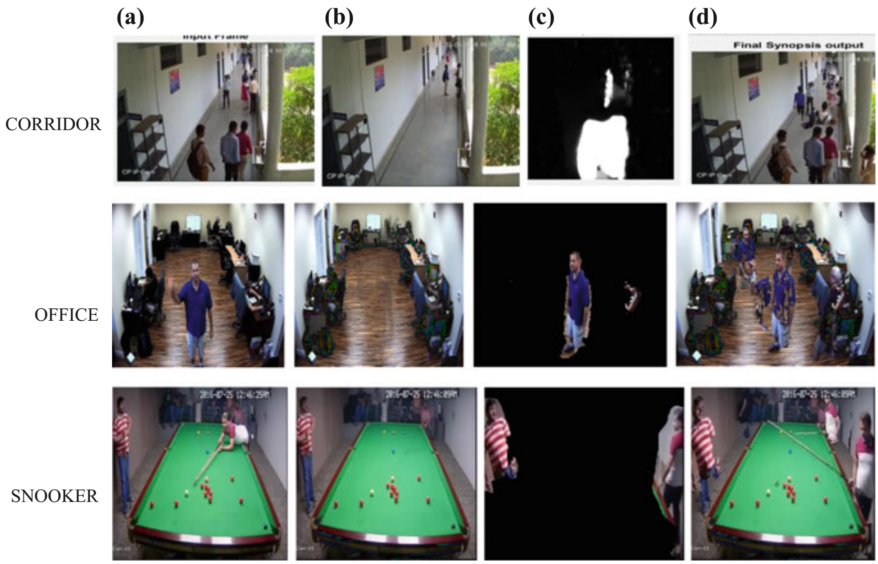


Fig. 8 Process flow of video synopsis; **a** The original video **b** Extracted background **c** Extracted foreground **d** final result video synopsis

4 Conclusion

In this paper, an effective method of video rundown is proposed. In that, the input video can be converted into frames and stored in the program. This method can utilize GUI commands from user and produce digested video. In this technique, dynamic indexing is used to track the moving objects. A facility is provided to the users for selecting keyframes in which they interested.

Acknowledgements We need to thank the accommodating remarks and recommendations from the unknown analysts. The proposed algorithm is developed by me and images which are utilized in this work are taken with the help of CCTV cameras, except office, and snooker videos. These are downloaded from the Google and open source data set.

References

1. Rubinstein M, Shamir A, Avidan S (2008) Improved seam carving for video retargeting. ACM Trans Graph 27(3):16. ACM
2. Zhong R, Hu R, Wang Z, Wang S (2014) Fast synopsis for moving objects using 925 compressed video. Signal Process Lett IEEE 21(7):834–838
3. Stauffer C, Grison WEL (1999) Adaptive background mixture models for real-time tracking. In: CVPR, pp 246–252

4. Boult TE, Micheals RJ, Gao X, Eckmann M (2001) Into the woods: visual surveillance of noncooperative and camouflaged targets in complex outdoor settings. Proc IEEE 89:1382–1401
5. Zhang Z, Huang K, Tan T (2008) Multi-thread parsing for recognizing complex events in videos. In: Torr P, Zisserman A (eds) 10th ECCV, Part III, pp 738–751
6. Zeng W, Du J, Gao W, Huang Q (2005) Robust moving object segmentation on H.264/AVC compressed video using the block-based MRF model. Real-Time Imaging 11:290–299
7. O'callaghan D, Lew EL (1995) Method and apparatus for video on demand with fast forward, reverse and channel pause, US Patent 5,477,263, 19 Dec 1995
8. Gandhi NM, Misra R (2015) Performance comparison of parallel graph coloring algorithms on bsp model using hadoop. In: International conference on computing, networking and communications (ICNC). IEEE, pp 110–116
9. Zhong R, Hu R, Wang Z, Wang S (2014) Fast synopsis for moving objects using compressed video. IEEE Signal Process Lett 21(7):834–838
10. Li Z, Ishwar P, Konrad J (2009) Video condensation by ribbon carving. IEEE Trans Image Process 18(11):2572–2583
11. Yoo JW, Yea S, Park IK (2013) Content-driven retargeting of stereoscopic images. IEEE Signal Process Lett 20(5):519–522
12. Panagiotakis C, Ovsepian N, Michael E (2013) Video synopsis based on a sequential distortion minimization method. In: International conference on computer analysis of images and patterns
13. Ye Y, Yi-jun L, Yan-qing W (2014) An improved aco algorithm for the bin packing problem with conflicts based on graph coloring model. In: International conference on management science & engineering (ICMSE). IEEE, pp 3–9
14. Pritch Y, Rav-Acha A, Peleg S (2008) Nonchronological video synopsis and indexing. IEEE Trans Patt Anal Mach Intell 30(11):1971–1984
15. Javed O, Shafique K, Shah M (2007) Automated visual surveillance in realistic scenarios. IEEE Multimedia 14:30–39
16. Babu RV, Ramakrishnan KR, Srinivasan SH (2004) Video object segmentation: a compressed domain approach. CSVT 14:462–474
17. Oh J, Wen Q, Hwang S, Lee J (2004) Video abstraction, video data management and information retrieval, pp 321–346
18. Yeung MM, Yeo B-L (1997) Video visualization for compact presentation and 855 fast browsing of pictorial content. Circuits Syst Video Technol IEEE Trans 7(5):771–785
19. Huang C-R, Chung P-CJ, Yang D-K, Chen H-C, Huang G-J (2014) Maximum a posteriori probability estimation for online surveillance video synopsis. Circuits Syst Video Technol IEEE Trans 24(8):1417–1429
20. Feng S, Lei Z, Yi D, Li SZ (2012) Online content-aware video condensation. In: IEEE Conference 930 on computer vision and pattern recognition (CVPR). IEEE, pp 2082–2087

An Intrusion Detection and Prevention System Using AIS—An NK Cell-Based Approach

B. J. Bejoy and S. Janakiraman

Abstract The widespread use of internet in key areas has increased unauthorized attacks in the network. Intrusion detection and prevention system detects as well as prevents the attacks on confidentiality, integrity, and availability of the system. In this paper, an Artificial Immune System based intrusion detection and prevention system is designed using artificial Natural Killer (NK) cells. Random NK cells are generated and negative selection algorithm is applied to eliminate self-identifying cells. These cells detect attacks on the network. High health value cells that detect a large number of attacks are proliferated into the network. When the proliferation reaches a threshold, the NK cells are migrated into the intrusion prevention system. So NK cells in IDS are in promiscuous mode and NK cells in IPS are in inline mode. The technique yields high detection rate, better accuracy and low response time.

1 Introduction

Internet and its fabulous growth have constituted to the association of a large number of people around the world. Due to its fast evolution, all basic services now depend on the internet. This leads to a growth in attacks on basic services by malicious users. Intrusion detection and prevention system (IDPS) is used to preserve the reliability, privacy, and accessibility of the system. Any malicious activity conveyed is either notified to the administrator or some systems even have the ability to respond to an intrusion. The systems that respond to an intrusion is acknowledged as intrusion prevention system. This is done by dropping the packets or blacklisting the IP address used by the invader. Anderson [1] developed an IDS by audit trail analysis in 1980 and Denning [2] established a model that become the base of modern IDS by anomaly detection method of both user and system data.

B. J. Bejoy (✉) · S. Janakiraman
Department of Banking Technology, Pondicherry University, Pondicherry, India
e-mail: bejoybj@gmail.com

S. Janakiraman
e-mail: jana3376@yahoo.co.in

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_86

The Human immune system (HIS) is a perfect defense mechanism that fights against diseases without any past information about these diseases. These mechanisms were impeccably similar to intrusion detection and prevention systems in the computer. The adaptation of human immune system's accomplishments to computer security emerged as a field branded as Artificial Immune System (AIS). IDPS based on AIS are generally anomaly detection models [3, 4]. In anomaly detection, a system using antigens (non-self) is fabricated using self-antigen (normal) and what the system recognizes is identified as an attack.

2 Related Work

AIS models were very effectual in designing an anomaly based IDPS system. Amalgamation of agent-based models with AIS developed a frightful combination for impostors attacking a system. AIS designs are based on four AIS Algorithms. The foremost concept is a Negative selection [4] which is commonly used to eliminate self-identifying agents. The next concept is Clonal Selection [5, 6] that is used for cloning of high fitness value agents. The third is the Danger theory [7, 8] used as a filtering technique in anomaly detection. Fourth is the Immune network [9] used mainly for communication between agents.

A dynamic real-time system was proposed by means of immune network algorithm was used for intrusion detection known as DIDAIN [10]. In this approach, three phases are used for intrusion detection. In the first phase, detectors are randomly generated and negative selection algorithm (NSA) is applied to eliminate self-identifying cell. Remaining cells are evolved into mature. During the second phase, the mature detectors are dynamically updated. When a mature detector detects another detector it is stimulated and when it is detected by another detector it is suppressed. The original intrusion detection happens in the third phase where the mature detectors are used for detecting any abnormal data. When the detector detects any data i.e. the data falls within the radius of a mature detector, the data is classified as abnormal. A multi-agent distributed intrusion detection system [11] was proposed which was used in virtual machines (VM). The system used negative selection algorithm for eliminating randomly selected detectors that identify self. During the training phase, the detectors with high fitness values were matured and migrated to the virtual machines. Clonal selection algorithm was used for cloning detectors with highest fitness value i.e. detectors that identify a large number of attacks have a high fitness value. Communication between agents takes place using immune network algorithm. The approach used static agents as well as mobile agents that travel to Virtual machines. Orchestra was used for agent communication and KVM hypervisor was used for implementing VM.

The concept of using Natural Killer Cells for Host-based intrusion detection was first devised in Fu et al. [12]. NK cells were used to detect concealed spyware in the system. If any infiltration behavior was exhibited by the program, activating signals were activated and when normal programs encounter NK cell, the inhibitory signal was activated. Low fitness NK cells were eliminated and higher fitness value cells were adapted. It even induces baits for concealed spyware to activate itself by presenting fake activities of the user. Mutation and crossover of NK cells were performed according to the fitness value of the NK cell. An Agent-based model was proposed known as ABIDS [13] that used dendritic cell algorithm (DCA). DCA is based on danger theory of HIS. The work used four types of agents. An antigen agent that sends a message to DC agent when it encounters a non-self antigen while parsing the dataset. A dendritic cell agent is distributed at the hosts that evaluate the danger value of antigens. T-Cell agents are activated when danger value exceeds a certain threshold and they inform the responding agent which response either by informing the security manager or by preventing the intrusion.

A real-time distributed network intrusion detection was proposed in Yang et al. [14] known as DAMIDAIS. Here multiple agents were used that interact with each other. Sensor agents were used to monitoring the environment. Analyzer agents were used for accumulating the information obtained from sensor agents. A manager agent is there to control the entire system. Message agents to carry the message between sensor agents. An alert agent was used to keep track of alert messages. An immune cooperation based learning (ICL) [15] was proposed based on immune cooperation mechanism. In this approach, the danger zone is not defined, but two signals were used. An antigen-specific signal from malware and an antigen-nonspecific signal from normal programs. These signals are taken together and the system is trained based on the cooperation of these two signals. A cooperative immunology based approach [16] that combines danger theory and Self/Non-Self theory for ids is presented. A sniffer module was used to capture the packets. A Non-Self Detector Module was used that compares the self-set with the captured packets and affinity was reported. A Danger detection module was used for comparing normal system profile with current profile. A vaccination Module was used that updates the system knowledge database. A result was produced by the decision module and response module was used for responses.

A Multi operational (MO) algorithm that used immune theory for fault detection was implemented [17] using Self/Non-Self theory. In this approach, random detectors were generated with variable detection radius for better coverage. For overlapped detectors, a moving detector algorithm was used. The detectors moved until the overlap measurement was less. If it does not reduce the overlap the detector radius decay was used. It was applied for fault detection in DC motor. Heavyweight agents, as well as various lightweight agents, were used for intrusion detection was proposed in Janakiraman and Vasudevan [18]. Even though the heavyweight agent increased the complexity of the system, it also increased the accuracy of the system. Ant colony optimization was used to detect intrusion in this approach. A detailed survey of AIS based IDS is presented in Bejoy and Janakiraman [19].

Even though there exists a large number of works using AIS, an effective multi-agent AIS based IDPS is still a deception and also many works in the survey lack an important parameter used for IDPS i.e. response time. These limitations motivated us to design an IDPS based on AIS using Natural Killer Cells.

3 Proposed Work

An Intrusion detection and prevention system the centered on Natural Killer (NK) Cells of AIS is projected in this paper. In the first section, we are giving an introduction to Natural Killer Cells and how artificial NK cells are used to design an IDPS. We propose an anomaly based IDPS system that can detect as well as respond to an intrusion. First, we design an IDS using NK cells and high health NK cells then migrate into the IPS thus reducing the probability of normal packets to be dropped. Random NK cells are created and the negative selection algorithm is applied to eliminate self-identifying cells. Then high health cells are cloned and based on IP threshold, effective cells are migrated into the IPS that are inline with the traffic.

3.1 Natural Killer Cells

Natural Killer (NK) Cells are lymphocytes (white blood cells) of Innate Immune System that has the capacity to detect both viral infected cells and tumors. They are cytotoxic in nature that contains special proteins that perform apoptosis or programmable cell death. NK cell activation is mainly based on a molecule known as major histocompatibility complex class 1 (MHC1) present on target cell surface. If the MHC1 class is missing NK cells perform lysis or apoptosis, hence they are called natural killers. Once considered the mainstay of innate immunity, recent studies [20] suggest that they have immunological memory also.

3.2 Artificial NK Cells

An artificial NK cell is defined as “An autonomous AIS cell that proliferates into the network and high health cells are then migrated into the Intrusion Prevention System”. In this paper, we are considering NK cells with immunological memory for designing IDPS. NK cells have two type of receptors-an Activating Receptor(AR) and an Inhibitory Receptor (IR). AR is activated when an NK cell detects an antigen and an IR is activated when it encounters a normal cell.

An Artificial NK cell with immunological memory can be expressed as follows

NK (Type, AR, Health, State, CN)

Type: Types of attacks an NK cell is trained to detect. An NK cell can detect multiple numbers of attacks.

Activating Range (Ar): The range from where an NK cell can detect an attack. If the incoming MHC1 falls in the range of AR, then the NK cell detects the attack.

Health: Health value indicates the fitness value of an NK cell. When an NK cell detects an attack, its health value increases.

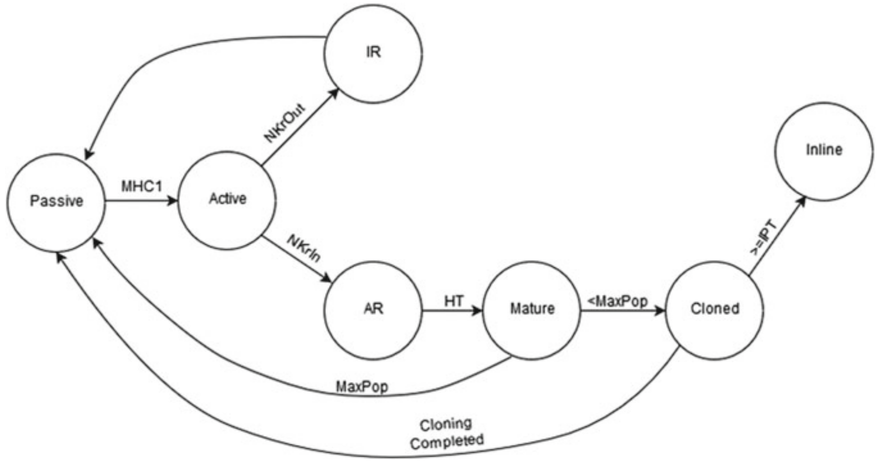


Fig. 1 State transition diagram of an NK cell

State: State indicates the states of an NK cell. There are seven states of an NK cell that are explained later. The initial state is the passive state.

CN: Number of clones indicates the total number of cloned NK cell in the network. This is the main eligibility of an NK cell to be used in the Intrusion Prevention system deployed in the network. When CN exceeds the Intrusion Prevention threshold the NK cell is migrated into the Intrusion prevention system.

An NK cell in this paper has seven states. Initially, an NK cell is in a passive state. When it encounters an MHC1 value, it changes to active state. If the MHC1 value is inside the NK cell radius (NKr), then the NK cell changes to Activating response (AR) state. If the incoming MHC1 is outside the NKr then the NK cell is in Inhibitory Response state (IR). MHC1 outside NKr means that it is a normal packet and hence no response from the NK cell. So NK cell returns back to the Passive state. On the other hand, if it's in AR means that the Nk cell detected an attack. Hence the health value is incremented. If the health value reaches a particular threshold (HT) the NK cells are in the Mature state. When NK cell is in the mature state the system checks whether the maximum population is reached (MaxPop). If MaxPop is reached the NK cell returns to the passive state. If MaxPop is not reached, NK cell is proliferated and it changes to Cloned State. When cloning is completed, the number of total clones is calculated and compared with Intrusion Prevention Threshold (IPT). If its less than the IPT, the NK cell returns back to the Passive state. If Number of clones is greater than IPT the cell is in inline state and migrated to the intrusion prevention system. After reaching IPT an NK is not proliferated hereafter. The state transition is shown in Fig. 1 and states are described in Table 1.

Table 1 States description

States	Description
Passive	Initial state
Active	When it encounters MHC1
IR	When MHC1 inside NKr
AR	When MHC1 outside NKr
Mature	When health > Health threshold
Cloned	When number < Maximum population
Inline	When number of clones >= Intru.Preven.Threshold

3.3 Working

During the initial stage, random NK cells are created based on Eq. (1).

$$\text{NKR}_i = \text{Min} \sum_{j=1}^{M} \left(C_i - R_{ij}\right)^2 \tag{1}$$

where $\text{NK}_i = (C_1, C_2 \ldots)$ and $S_j = (R_{1j}, R_{2j} \ldots)$.

S_j refers to the self -antigens and NK_i refer to NK cells and C and R are the detectors and self-antigen coordinates respectively. More than one attacks can be identified by an NK cell. This helps us to eliminate self-identifying cells and thus negative selection algorithm is used to eliminate such cells. Throughout testing phase, when an NK cell detects an attack its health value increases. When the health value reaches a threshold an NK cell is proliferated.

3.4 Working of NK Cell

MHC1 is created based on the incoming packets. MHC1 can be extracted from port number, IP address etc. MHC1 is given to NK cell. If MHC1 matches any detectors, then activating response is activated. It means the incoming MHC1 is an attack. If it does not match any detectors, then the inhibitory response is activated and MHC1 is considered normal. The working of NK cell is illustrated in Fig. 2.

3.5 NK Cell-Based IDPS Architecture

NK cell-based IDPS architecture consists of Sensors, Analyzers, IDS module and IPS module. Passive sensors are used which collect a copy of the incoming traffic and provide it to analyzers as NK cells are positioned in promiscuous mode. When

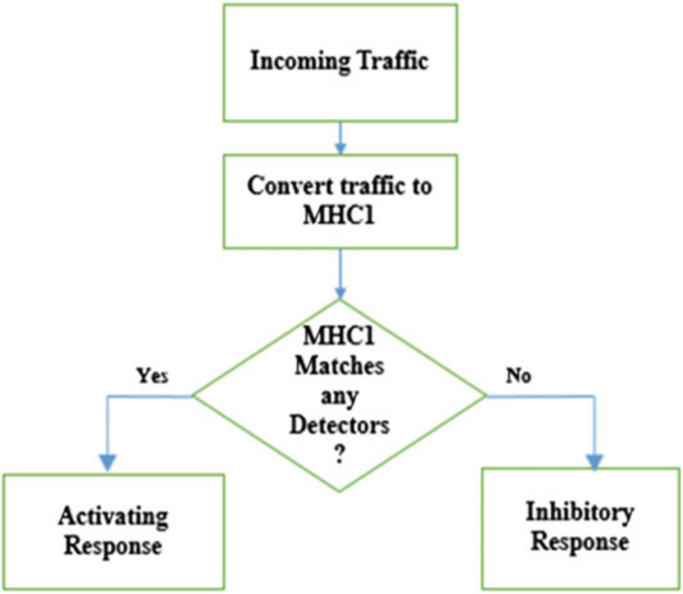


Fig. 2 Working of NK Cell

NK cells detect various attacks the health value is increased which result in the proliferation of NK cells. Analyzer extracts only useful information from the packet that is used to create an MHC1. The analyzer also accomplishes load balancing when a number of NK cells are available. All NK cells work parallelly so that the overall detection time is very quick. When the number of cloned NK cell reaches a threshold value, the NK cell migrates into the Intrusion Prevention System module that connects the NK cell to the incoming traffic in inline mode. NK cell-based IDPS is illustrated in Fig. 3.

4 Results and Discussion

We implement the work on MATLAB R2017a. 10% of NSL KDD Cup 99 dataset [21] is used as it is the only popular benchmark dataset available for IDS. The dataset is preprocessed to eliminate redundant entries and normalized before giving to the algorithm. Min–max normalization is used to make the data within a range from 0 to 1. The output was verified based on the five classes available in the dataset—Probe, DoS, U2R, R2L and Normal. The work was compared with that of [17, 18]. Four parameters were used to evaluate the work. Detection rate (DR), Accuracy (Acc), False alarm rate (FAR) and Response Time (RT) [19].

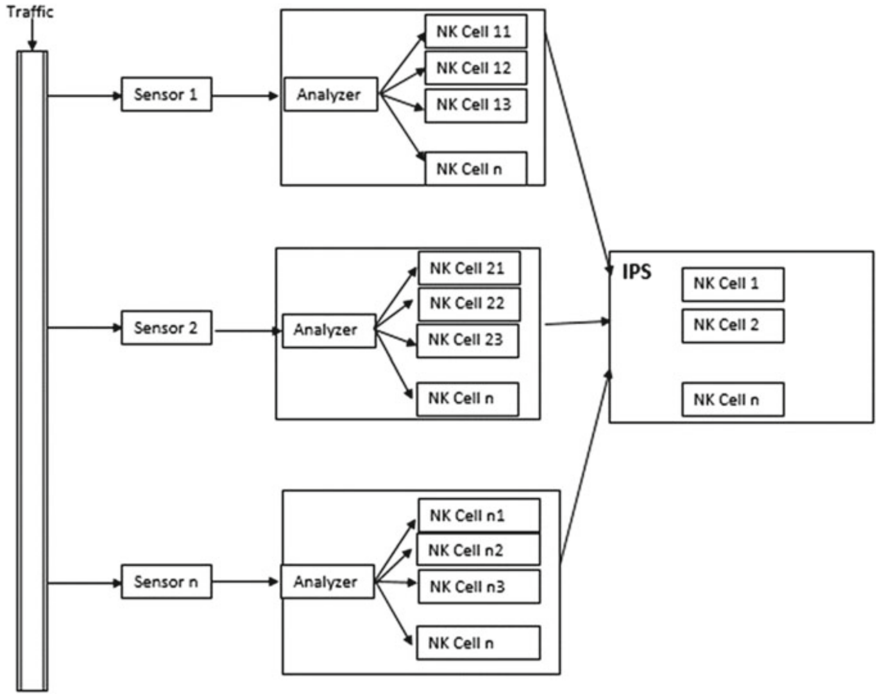


Fig. 3 NK Cell-based IDPS architecture

$$\mathrm{DR} = \mathrm{TP}/(\mathrm{TP} + \mathrm{FN}) * 100 \tag{2}$$

$$\mathrm{Acc} = (\mathrm{TP} + \mathrm{TN})/(\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}) * 100 \tag{3}$$

$$\mathrm{FAR} = \mathrm{FP}/(\mathrm{FP} + \mathrm{TN}) * 100 \tag{4}$$

$$\mathrm{RT} = \mathrm{Dd} + \mathrm{Dr} \tag{5}$$

where TP-True Positive, FN-False Negative, TN-True Negative, FP-False Positive, Dd-Detection Delay and Dr-Response delay.

NK cell-based approach has an average detection rate of 96.7% whereas the other two approaches have 95.6 and 95% respectively (Fig. 4).

The average accuracy of (Fig. 5) NK cell-based approach is 99.4% whereas the other two approaches have 99.1 and 98.4% respectively.

The average FAR for NK cell-based approach is 0.4 while others have 0.7 and 1.1 respectively (Fig. 6). Low FAR of NK cell approach makes it appropriate to be used for Intrusion prevention system.

The NK cell-based approach has a quick response time than other two approaches making it an idle candidate for IPS (Fig. 7).

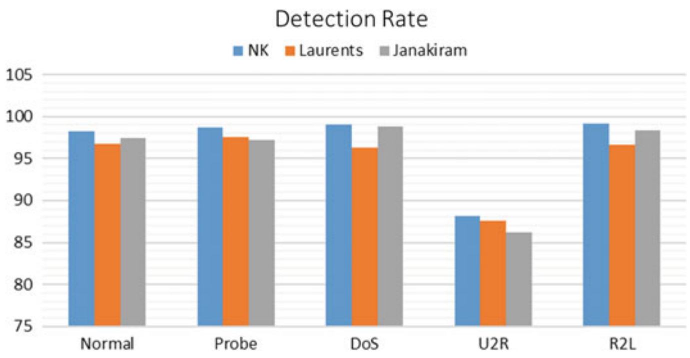


Fig. 4 Detection rate

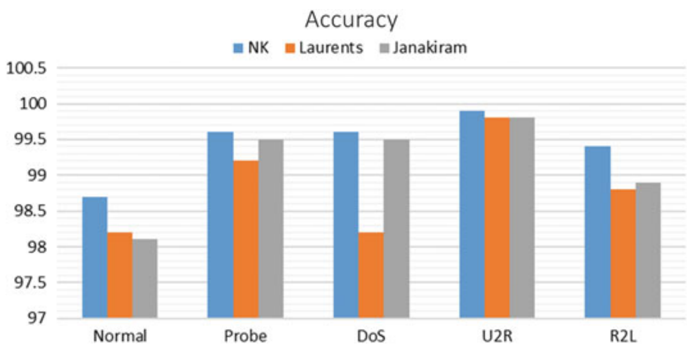


Fig. 5 Accuracy

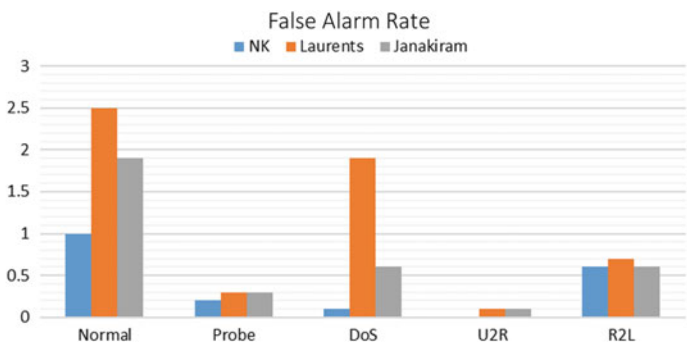


Fig. 6 False alarm rate

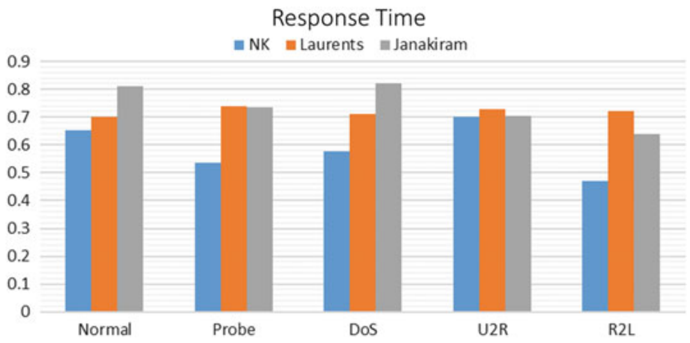


Fig. 7 Response time

5 Conclusion and Future Work

An NK cell-based Intrusion detection and prevention system is suggested. The approach has high accuracy, better detection rate, low false alarm rate and quick response time making it idle to be used in intrusion detection and prevention systems. Random detectors are generated and negative selection is used to eliminate self-identifying detectors. High health NK cells are proliferated using the clonal selection algorithm. Frequently cloned NK cells are migrated into an intrusion prevention system that can give an active response by dropping packets.

Future work includes using a Heavyweight agent that can detect new attacks as well as improve the accuracy. The system will be tested real time in a network.

References

1. Anderson JP (1980) Computer security threat monitoring and surveillance. Washing, PA, James P. Anderson Co
2. Denning DE (1986) An intrusion detection model. In: Proceedings of the seventh IEEE symposium on security and privacy
3. Forrest S, Hofmeyr SA, Somayaji A (1997) Computer Immunology. Commun ACM 40(10):88–96
4. Forrest S, Pereslon AS, Allen L, Cherukuri R (1994) Self-nonself discrimination in a computer. In: Proceedings of the 1992 IEEE symposium on security and privacy. IEEE Computer Society Press, pp 202–212
5. Burnet FM (1959) The clonal selection theory of acquired immunity. Vanderbilt University Press, USA
6. de Castro LN, Von Zuben FJ (2002) Learning and optimization using the clonal selection principle. IEEE Trans Evol Comput 6:239–251
7. Matzinger P (2002) The danger model: a renewed sense of self. Science 296(5566):301–305
8. Greensmith J, Aickelin U, Cayzer S (2005) Introducing dendritic cells as a novel Immune-Inspired algorithm for anomaly detection. Lecture Notes in Computer science, 3627. Springer, Berlin, Heidelberg

9. Jerne NK (1974) Towards a network theory of the immune system. Ann Immunol (Inst Pasteur) 125C:373–389
10. Hu X, Liu X, Li T, Yang T, Chen W, Liu Z (2015) Dynamically real-time intrusion detection algorithm with immune network. J Comput Inf Syst 11:587–594
11. Afzali Seresht N, Azmi R (2014) MAIS-IDS: a distributed intrusion detection system using multi-agent AIS approach. Eng Appl Artif Intell 35:286–298
12. Fu J, Yang H, Liang Y, Tan C (2012) Bait a trap: introducing natural killer cells to artificial immune system for spyware detection. Lecture Notes Computer Science (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), 7597. LNCS, pp 125–138
13. Ou CM (2012) Host-based intrusion detection systems adapted from agent-based artificial immune systems. Neurocomputing 88:78–86
14. Yang J, Liu X, Li T, Liang G, Liu S (2009) Distributed agents model for intrusion detection based on AIS. Knowl-Based Syst 22:115–119
15. Zhang P, Tan Y (2015) Immune cooperation mechanism based learning framework. Neurocomputing 148:158–166
16. Sobh TS, Mostafa WM (2011) A cooperative immunological approach for detecting network anomaly. Appl Soft Comput J 11:1275–1283
17. Laurentys CA, Ronacher G, Palhares RM, Caminhas WM (2010) Design of an artificial immune system for fault detection: a negative selection approach. Expert Syst Appl 37:5507–5513
18. Janakiraman S, Vasudevan V (2009) Agent-based DIDS: a intelligent learning approach. Int J Intell Inf Process 8. Serials Publications
19. Bejoy BJ, Janakiraman S (2017) Artificial immune system based intrusion detection systems—a comprehensive review. Int J Comput Eng Technol 8(1):85–95
20. Arina A, Murillo O, Dubrot J, Azpilikueta A, Alfaro C, PÃrez-Gracia JL, Bendandi M, Palencia B, Hervás-Stubbs S, Melero I (2007) Cellular liaisons of natural killer lymphocytes in immunology and immunotherapy of cancer. Expert Opin Biol Ther 7(5):599–615
21. NSL KDD CUP 99 Dataset. http://nsl.cs.unb.ca/NSL-KDD

HORBoVF—A Novel Three-Level Image Classifier Using Histogram, ORB and Dynamic Bag of Visual Features

Vishwas Raval and Apurva Shah

Abstract Every country has its own currency in terms of coins and paper notes. Each of the currency of Individual County has its unique features, colors, denominations, and international value. Though it is easy for us to identify the denomination but for the blind people, it is a not at all possible to do this! Especially when size of currency of different denominations is same, it becomes almost impossible for them to do this and there are chances that they might got cheated by others. This paper proposes and discusses a novel three-stage image classifier algorithm for the same purpose to help the blind people in identifying denomination more accurately and to check if the currency is real or fake.

1 History of Currency

In early age of human civilization, people used to have barter systems to carry out daily routines (transactions!). Gradually, barter system became obsolete and currencies came into existence for day to day life from 2000 BC where it used to be in the form of coins. Paper-based currencies started in between 618 AD and 907 AD, in pre-modern china, during Tang dynasty. During seventh to twelfth century, the paper-based currency was introduced in the Islamic countries which later on became the base for a stable and a high valued currency called “dinar”. In 1661, it was Sweden which introduced the first paper-based currency in Europe. Each country in this world has its own currency with specific denomination indicating its monetary value. US Dollars, British Pound, Japanese Yen, EURO are the examples of currencies of different countries.

V. Raval (✉) · A. Shah
CSE Department, Faculty of Technology and Engineering,
The M S University of Baroda, Vadodara, India
e-mail: vishwas.raval@gmail.com

A. Shah
e-mail: apurva.shah-cse@msubaroda.ac.in

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_87

In India, Rupee is the main currency and printing and distribution is controlled by the Reserve Bank of India. Till 2010, “Rs.” was the word which was used for Indian currency which was replaced by the symbol for ₹, designed by D. Uday Kumar, and Government of India (GoI) conferred it as an official symbol for Rupee.

Traditionally, starting from British rule, Indian currency was ranging from 1 Aana to 100 Rupees. The currency had denomination like 1, 5, 10, 20, 25, and 50 in the form of Paise. Later on, new coins of ₹1, ₹2, ₹5, ₹10 were introduced and rests of the coins have been taken back from market. The currency notes are available in a denomination value of ₹5, ₹10, ₹20, ₹50, ₹100, ₹200, ₹500, and ₹2000.

When, on November 8, 2016, Hon’ble Prime Minister of India demonetized the existing currencies of ₹500 and ₹1000 and introduced new ₹500 and ₹2000 currencies. In his speech, he told the motive behind demonetization, to curb the black money, eradiate corruption menace and prevent the terror funding and Hawala business being carried out for terrorist activities in J&K.

2 Motivation, Existing Tools, Applications and Related Work

Money is something for which people gets cheated in different ways. When it comes to blind people, it becomes nightmare for them to identify it and perform daily transactions. Talking with them, they said that they rely on others and what others say, they have to trust. In addition to this, post-demonetization, GoI introduced few new currencies of ₹200, ₹500, and ₹2,000. The dimensions of these new currencies are, in fact, smaller than regular currency notes. This has made blind people’s life more miserable. Earlier, the size of ₹2,000 was the largest one, then ₹500, then ₹100 and so on. Now, the ₹500 currency is smaller in size than ₹100. So there are chances that blind people might get cheated as they do not have now any way out to identify it. Though GoI has introduced some embossing for blind people but that embossing becomes useless as the currency gets older [1]. So, the only motivation behind this work is to help the blind people in currency identification, where there are highest chances of being cheated.

In United States, all the currency denominations are of same size which makes the identification more difficult for blind people. In such situation, the Governments provide some way to help them to identify the different money denominations. The countries like Australia and Malaysia have distinct sized notes for various denominations which helps the blind people to recognize the denomination easily. In Canadian Dollars, for every denomination, there is a specific Braille mark so that the blind people can easily read those Braille marks and recognize the denomination.

Smart Saudi Currency Recognizer (SSCR) [2] is a currency recognizer for Saudi Arabia. LookTel App [3] supports the US Dollar, Australian Dollar, Bahraini Dinar, Brazilian Real, Belarusian Ruble, British Pound, Canadian Dollar, Euro, Hungarian Forint, Israeli Shekel, Indian Rupee, Japanese Yen, Kuwaiti Dinar, Mexican Peso, New Zealand Dollar, Polish Zloty, Russian Ruble, Saudi Arabian Riyal, Singapore Dollar, and United Arab Emirates Dirham. EyeNote [4] and MoneySpeaker [5] are the other applications used for currency recognition.

Apart from these applications, many research oriented attempts, in the form of algorithms, have been carried out in order to give a robust way for currency recognition across the world since 1992. For all the work that has been carried out, the preprocessing and feature extraction are the common techniques which have been used in the initial phase. The following is the list of the classification approaches that have been used in the individual work in order to recognize the currency accurately (Table 1).

In India, the currency recognition tools are available with ATMs and Banks. But these tools are not affordable to everyone cannot afford and not handy as well. Hence, there is a real need for a system that can help the blind people in India to recognize currency properly, especially in the situations like demonetization when real currency also becomes un-useful.

3 HORBoVF—Histogram and ORB Based Bag of Visual Features Classifier

Seeing the demand of a robust, affordable and handy system to help the visually challenged people in currency identification, the authors have developed a computationally lighter computer vision algorithm. This was earlier named as ***iCu₹e*** [6], now ***D₹ushti***, an Android App for Indian currency recognition in Indian vernacular languages. The main reason behind having an Android App is the blind people are quite acquainted in using smartphones with Talk-Back feature. The algorithm has been developed in order to provide accuracy in currency identification and fake currency recognition. At the same time, as it has been targeted for the mobile devices so care has been taken to make it faster and lighter so as to run on low memory devices too.

The overall algorithm has been divided into three stages after Image Capture, called Preprocessing, Image Classification (Identification) and Text-to-Speech Conversion. The preprocessing performs background removal in order to improve accuracy and remove outliers and irrelevant parts from the background using GrabCut [7]. The Image classification itself has been divided into three stage Filtered Classification process.

First, the captured image is converted into histogram and matched with dataset. Histogram intersection is used for finding the closest match. Here, top 10 histogram intersections are fetched. However, the histogram of a banana and a grassfield could have more intersection value and may lead to wrong identification. To filter-out any such Banana-Grassfield combination, the images with top ten histogram intersections are passed to the next layer of classification. At second layer, Oriented BRIEF (ORB) [8], alternative to SIFT and SURF, is used for feature extraction and matching. The

Table 1 List of currency recognition algorithms/work carried out for various currencies

Sr.#	Currency recognition approach	Currency	Accuracy	Year
1	Neural network [11]	USD	98.08	1993
2	Neural network, optimized mask and genetic algorithm [12]	USD	>95%	1995
3	Hybrid neural network [13]	USD & Japanese	92	1996
4	Multilayered peceptrons in NN [14]	USD		1996
5	Neural NN with Gaussian distribution [15]	USD, CA$, AU$, Krone, Franc, GBP, Mark, Pesetas		1998
6	Neural NN and axis-symmetrical mask [16]	EURO	97	2000
7	Neural NN and principal component analysis [17]	USD	95	2002
8	Back propagation NN [18]	Chinese Renminbi	96.6	2003
9	Speeded-up robust feature (SURF) [19]	USD		2007
10	Markov models [20]	USD, EURO, Dirham, Rial	95	2007
11	Ada-boost classification [21]	USD		2008
12	Artificial neural network [22]	SL Rupee		2008
13	Neural network [23]	Malaysian Ringit		2008
14	Data acquisition [24]	Chinese Renminbi	100	2008
15	Bio-inspired image processing [25]	EURO	100	2009
16	Ensemble neural network with negative correlation learning [26]	Bangladeshi Taka	98	2010
17	Local binary patterns [27]	Chinese Renminbi	100	2010
18	Wavelet transform [28]	Rials	81	2010
19	Intersection change [29]	Chinese Renminbi	97.5	2010
20	Image processing & neural network [30]	Indian Rupee		2010
21	Support vector machine [31]	Chinese Renminbi	87.097	2011
22	Speeded-up robust feature (SURF) [32]	USD	100	2012
23	Wavelet transform & neural network [33]	Dirham	99.12	2012
24	Local binary patterns and RGB space [34]	Mexican	97.5	2012

(continued)

Table 1 (continued)

Sr.#	Currency recognition approach	Currency	Accuracy	Year
25	Quaternion wavelet transform & generalized Gaussian density [35]	USD, Renminbi and EURO	99.68	2013
26	Basic feature extraction using Euler numbers [36]	Pakistani Rupee		2013
27	Number recognition [37]	Chinese Renminbi	95.92	2014
28	Instance retrieval and indexing [38]	Indian Rupee	96.7	2014
29	Radial basis Kenrel function [39]	Dirham	91.51	2015
30	Segmentation, feature extraction [40]	Bangladeshi Taka		2015
31	Region of interest (ROI), discrete wavelet transform, linear regression and SVM [41]	Indian Rupee		2015

best matched image passing through a specific threshold value is selected from the top ten images. This image is sent to the third and final stage of classification for verification purpose.

Bag of Words classification [9, 10] is one a preferred technique for image classification. Here, in the third stage, Bag of Visual Features is created for the dataset images. ORB is used to detect features and feature descriptor is computed for those images. These descriptors form a raw bag of features is created. These features are clustered using K-Means algorithm in order to create final visual vocabulary. The test image received from second stage is compared with visual vocabulary and appropriate descriptor is calculated to label the image and converted into speech using Tect2Speech. The *HORBoVF* structure is shown in Fig. 1.

The pseudo code of the algorithm is given below

1. Initialize *totalNoOfImages (N)* in the dataset, the *featureSet* with distinguishable features for all images, *imageObject*, in the dataset.
2. Initialize *featureThreshold, similarityDistanceThreshold*
3. Histogram Intersection Filtering:
   a. Generate histograms of all N images, *imageObject*, in dataset and histogram of *inputImage*
   b. For each histogram of *imageObject*, perform histogram intersection with a histogram of *inputImage*
   c. Find top K histogram intersection values and corresponding *imageObject* into *topKIntersects* and *topKImageObjects*, where $K < N$.
4. ORB based decision making:
   a. Create an ORB feature detector for *inputImage*
   b. For each *imageObject* in *topKImageObjects* with the corresponding *featureSet*, Perform feature matching:

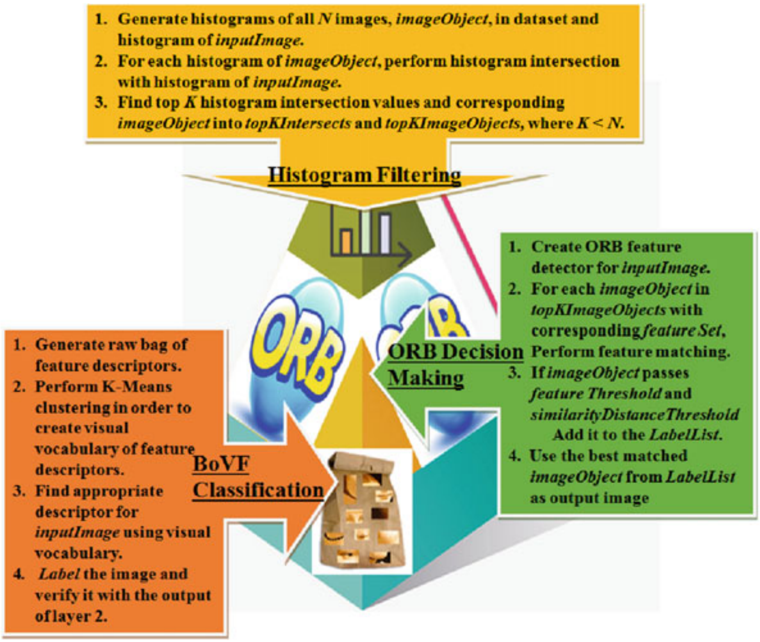


Fig. 1 The *HORBoVF* pyramid

If *imageObject* passes *featureThreshold* and *similarityDistanceThreshold* then,

Add it to the *LabelList*

c. Use the best-matched *imageObject* from *LabelList* as an output image.

5. Decision Verification using BoVF:

   a. Generate raw bag of feature descriptors.
   b. Perform K-Means clustering to create a visual vocabulary of feature descriptors.
   c. Find appropriate descriptor for *inputImage* using visual vocabulary.
   d. *Label* the image and verify it with the output of stage 4.

Note: *featureThreshold* and *similarityDistanceThreshold* are experimentally set values which may vary.

4 Testing and Conclusion

This paper proposes a three-layered image classifier for currency (image) recognition using Histogram, ORB, and Bag of Visual Features. The experiments are carried out on an Android (Kitkat) phone with 1 GB RAM and Quad-Core Max 1.6 GHz processor. Once the Visual Vocabulary is generated and loaded into the memory, which takes around 4–5 s, image classification is carried out in average 2–3 s with an average accuracy of 91.54% for $K = 24$ clusters. It has been observed also that wrong prediction occurs in case of improper capturing of image, how much portion of image covering distinguishable features is visible and mainly the distance of the object from camera.

Acknowledgements We are thankful to the Omnipotent God for making us able to do something for the society. We are thankful to our parents for bringing us on this beautiful planet. We are grateful to our department and University for providing support and resources for this work. Finally, we acknowledge the authors and researchers whose papers helped us to move ahead for this work.

References

1. http://indianexpress.com/article/india/visually-disabled-are-struggling-with-new-currency-notes-nab-demonetisation-note-ban-4621070
2. https://play.google.com/store/apps/details?id=sscr.imagemanipulations
3. http://www.looktel.com/moneyreader
4. http://www.eyenote.gov/
5. https://www.androidpit.com/app/com.hmi.moneyspeaker
6. Raval V, Shah A (2017) iCu₹e—an IoT application for indian currency recognition in vernacular languages for visually challenged people. In: Confluence-2017, 7th international conference on cloud computing, data science and engineering. Amity University, Noida
7. Rother C, Kolmogorov V, Blake A (2004) Grabcut: interactive foreground extraction using iterated graph cuts. ACM Trans Graph (TOG) 23(3):309–314. ACM
8. Ethan R, Vincent R, Kurt K, Gary B ORB: an efficient alternative to SIFT or SURF. In: IEEE international conference on computer vision
9. Cordelia S "Bag-of-features for image classification", Thesis, INRIA (2014)
10. Dimitri L (2006) Image classification with bags of local features. Ph.D. Thesis
11. Takeda F, Omatu S, Onami S (1993) Recognition system of US dollars using a neural network with random masks. IJCNN'93-Nagoya. In: Proceedings of 1993 IEEE international joint conference on neural networks
12. Takeda F, Omatu S (1995) A neuro-paper currency recognition method using optimized masks by genetic algorithm. In: IEEE International conference on intelligent systems, man and cybernetics for the 21st century, vol 5, pp 4367–4371
13. Takeda F, Sigeru O (1995) High speed paper currency recognition by neural networks. IEEE Trans Neural Netw 6:73–77
14. Frosini Angelo, Gori Marco, Priami Paolo (1996) A neural network-based model for paper currency recognition and verification. IEEE Trans Neural Netw 6:1482–1490
15. Masahiro T, Takeda F, Ohkouchi K, Michiyuki Y (1998) Recognition of paper currencies by hybrid neural network
16. Takeda F, Toshihiro N (2000) Multiple kinds of paper currency recognition using neural network and application for Euro currency. In: Proceedings of the IEEE international joint conference on neural networks, vol 2, pp 143–147
17. Ahmadi A, Sigeru O, Michifumi Y (2002) Implementing a reliable neuro-classifier for paper currency using PCA algorithm. In: Proceedings of the 41st SICE annual conference, vol 4, pp 2466–2468
18. Zhang E-H, Jiang B, Duan J-H, Bian Z-Z (2003) Research on paper currency recognition by neural networks. Int Conf Mach Learn Cybern 4:2193–2197

19. Chen W, Yingen X, Jiang G, Natasha G, Radek G (2007) Efficient extraction of robust image features on mobile devices. In: Proceedings of the 2007 6th IEEE and ACM international symposium on mixed and augmented reality, pp 1–2
20. Hassanpour H, Yaseri A, Ardeshiri G (2007) Feature extraction for paper currency recognition. In: 9th international symposium on signal processing and its applications, pp 1–4
21. Liu X (2008) A camera phone based currency reader for the visually impaired. In: Proceedings of the 10th international ACM SIGACCESS conference on Computers and accessibility, pp 305–306
22. Gunaratna DAKS, Kodikara ND, Premaratne HL (2008) ANN based currency recognition system using compressed gray scale and application for Sri Lankan currency notes-slcrec. In: Proceedings of World academy of science, engineering and technology, pp 235–240
23. Nurlaila H (2008) Currency recognition and converter system. Ph.D. dissertation. Universiti Malaysia Pahang
24. He J, Zhigang H, Pengcheng X, Ou J, Minfang P (2008) The design and implementation of an embedded paper currency characteristic data acquisition system. In: International conference on information and automation, pp 1021–1024
25. Parlouar R, Florian D, Marc M, Christophe J (2009) Assistive device for the blind based on object recognition: an application to identify currency bills. In: Proceedings of the 11th international ACM SIGACCESS conference on Computers and accessibility, pp 227–228
26. Debnath K, Sultan U, Shahjahan Md, Kazuyuki Mu (2010) A paper currency recognition system using negatively correlated neural network ensemble. J Multimedia 5(6):560–567
27. Guo J, Yanyun Z, Anni C (2010) A reliable method for paper currency recognition based on LBP. In: 2nd IEEE international conference on network infrastructure and digital content, pp 359–363
28. Daraee F, Saeed M (2010) Eroded money notes recognition using wavelet transform. In: 6th Iranian conference on machine vision and image processing, pp 1–5
29. Shao K, Yang G, Na W, Hong-Yan Z, Fei L, Wen-Cheng L (2010) Paper money number recognition based on intersection change. In: Third International workshop on advanced computational intelligence (IWACI), pp 533–536
30. Gopal K (2010) Image processing based feature extraction of Indian currency notes; MTech Thesis. Thapar University
31. Yeh C, Wen-Pin S, Shie-Jue L (2011) Employing multiple-kernel support vector machines for counterfeit banknote recognition. Appl Soft Comput 11(1):1439–1447
32. Hasanuzzaman F, Xiaodong Y, Yingli T (2012) Robust and effective component-based banknote recognition for the blind. IEEE Trans Syst Man Cybern (Appl Rev) 42(6):1021–1030
33. Ahangaryan F, Mohammadpour T, Kianisarkaleh A (2012) Persian banknote recognition using wavelet and neural network. In: International conference on computer science and electronics engineering (ICCSEE), pp 679–684
34. García-Lamont F, Jair C, Asdrúbal L (2012) Recognition of Mexican banknotes via their color and texture features. Expert Syst Appl 39(10):9651–9660
35. Gai S, Guowei Y, Minghua W (2013) Employing quaternion wavelet transform for banknote classification. Neurocomputing 118:171–178
36. Ali A, Mirfa M (2013) Recognition system for Pakistani paper currency. World Appl Sci J 28(12):2069–2075
37. Yu H, Yingyong Z (2014) Study on money number recognition arithmetic. Int J Multimedia Ubiquit Eng 9(11):189–196
38. Suriya S, Shushman C, Vishal K, Jawahar C (2014) Currency recognition on mobile phones. In: Proceedings of IEEE 22nd international conference on pattern recognition, pp 2661–2666
39. Sarfraz M (2015) An intelligent paper currency recognition system. In: Procedia Computer Science-65, pp 538–545
40. Saifullah S, Rahman M, Hossain Md (2015) Currency recognition using image processing. Am J Eng Res 4(11):26–32
41. Pham T, Danh Y, Seung Y, Dat T, Husan V, Kang R, Dae S, Sungsoo Y (2015) Recognizing banknote fitness with a visible light one dimensional line image sensor. Sensors 15(9):21016–21032

Neural Network Based Image Registration Using Synthetic Reference Image Rotation

S. Phandi and C. Shunmuga Velayutham

Abstract Typical image registration techniques use a set of features from a target and reference images and search in the affine transformation space using a similarity metric. Neural Networks typically have employed two choices—geometric transformations to find correlation between images and a similarity metric. In this paper, however, we have proposed and employed a simple and effective method for image registration using neural networks. The image registration has been formulated as a classification problem. By generating and learning exhaustive synthetic reference image transformations appropriate re-transformation for target image is computed for effective registration. The proposed work is tested on satellite imagery.

1 Introduction

Image Registration is the process of aligning multiple images of same scene into a single integrated image with a common co-ordinate system. In essence, this spatial alignment of images essentially involves determination of appropriate geometrical transformation that aligns one image with reference to a particular image. Registration is often carried out primarily for information fusion—multiple images offer more comprehensive information than provided by individual images. Typically, registration is most often employed in medical [1, 2] and satellite imagery [3–5] for multi-view, multi-temporal and multi-modal analyses. In both application domains, factors related to image acquisition devices, images themselves and objects in images often make registration a hard problem.

Most image registration methods, proposed in literature, typically involves *feature space* (used for matching), *search space* (of geometrical transformations), *a*

S. Phandi · C. Shunmuga Velayutham (✉)
Department of Computer Science and Engineering, Amrita School of Engineering, Coimbatore, Amrita Vishwa Vidyapeetham, India
e-mail: cs_velayutham@cb.amrita.edu

S. Phandi
e-mail: cb.en.p2cvi16009@cb.students.amrita.edu

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_88

search strategy and *a similarity metric*. Extrinsic methods [6], surface methods [7], moments and principal axes methods [8], correlation-based methods [9], mutual information based methods [10], and wavelet-based methods [11] are various image registration methods to cite by way of few examples are some of the methods of registration in literature. Soft computing based methods [12] are relatively recent and have acquired prominence by virtue of their capability to handle uncertainty effectively. Among softcomputing techniques, Neural Networks have been one of the popular and effective techniques employed for image registration.

Against the iterative optimization procedures, Neural Networks possess the ability to learn the input-output relationship from a data set with no prior knowledge. Typical image registration techniques use a set of features from a target and reference images and search in the affine transformation space using a similarity metric. Neural Networks, in the literature, typically have employed two choices—geometric transformations to find correlation between images and a similarity metric.

In this paper, however, we have proposed and employed a simple and effective method for image registration using neural networks. By generating and learning exhaustive synthetic reference image transformations (restricted only to rotation in this work), appropriate re-transformation for target image is computed for effective registration. The proposed work is tested on satellite imagery.

This paper is organized as follows. Section 2 presents the related works. Section 3 details the proposed neural network based image registration method. Section 4 presents experimental design, simulation results and analysis. Finally, Sect. 5 concludes the paper.

2 Related Works

In case of registration, same co-ordinate points from reference image and target image is often determined with the help of common features (like edge information, edge outline [13]) that are present in both images. In simplest form, features extracted from distorted image are trained using back propagation algorithm [13]. Target image co-ordinates are then fed into the trained network. In [14], Radial Basis function neural network has been used (with lesser training time) to determine geometrical transformation. DCT (Discrete Cosine Transform) co-efficient and moments are extracted from reference image and are used to train the network.

In [15], fourier transform of both reference and target images are fed into a feed-forward network to estimate transformation parameter viz. translation, rotation, and magnification to register images. Feature-based image registration in [16], investigates the efficacy of Speeded Up Robust Feature (SURF), Scale Invariant Feature Transform (SIFT), Maximally Stable Extremal Regions (MSER) features for registering images. These features are invariant to zoom, noise, rotation, and illumination and are suitable for image registration. RANSAC algorithm is used for outlier removal and Recall as well as RMSE measures are used to estimate quality of transformation. In case of high resolution images, spatial information [17] is very essential for effective registration.

Texture analysis is employed to find texture boundaries for satellite image registration in [3]. GLCM (Gray-Level Co-Occurrence Matrix) is used for texture analysis of images. Pattern recognition has been used to estimate rate of changes in satellite images for registration [4]. In [5], neural network has been employed to classify pattern like buildings, farm land etc. in satellite images.

3 Proposed Work

Essentially, the problem of image registration can be reduced to finding the appropriate transformation between the target and reference images. The typical transformation model used for image registration is affine transformation since it is sufficiently general by virtue of its capability to handle translations, rotations and scaling. This affine transformation, for a general 3D case can be represented as

$$x' = A \cdot x + b,$$

where A (a 3×3 matrix) embeds the rotations and scaling, x, x' and b are three dimensional arrays and respectively represent the original position of feature vectors, the transformed positions and the translation information.

Typical image registration techniques use a set of features from a target and reference images and search in the affine transformation space using a similarity metric. Neural Networks are by no means an exception to this. This demands that image registration needs two choices viz. the geometric transformations to be considered for finding correlation between images and a suitable measure of match. The proposed method, in this paper, formulates finding appropriate affine transformation as a classification problem. By generating and learning exhaustive reference image transformations (restricted only to rotation in this work), appropriate re-transformation for target image is computed for effective registration. This alleviates the need for a similarity metric which often has a considerable impact on the registration process. Algorithm for the proposed method is explained below

1. SURF features from Reference image is taken at each degree of rotation.
2. SURF features from Target image is taken without rotation of image.
3. Train the reference image features using Patternet architecture.
4. Match the Trained reference image features with target image features.
5. Find Magnitude and Direction between matched Reference and Target image features.
6. Find angle of rotation and register the image.

The reference image is rotated a full circle, say with 1° each, thus obtaining 360 images. From each of those rotated images, N features are extracted. This $360 \times N$ matrix forms the feature matrix that forms input to the neural network. Each rotated image represents a single class, thus going with our example we essentially have 360 classes! So, 360 × 1 class labels forms the teaching signal. The feature matrix along with the class label vector forms the training set for the neural network.

The neural network has been trained on the above said training set about different classes of rotated images. After having learnt the rotation classes, features are extracted from the target image without any rotation. This feature vector (of size $1 \times N$) serves as a test vector which when administered into the neural network will get classified. Then with the features from trained reference image and target images features are matched then into an appropriate rotation class. The rotation class, thus identified, can be considered as the rotation the target image has undergone during the imaging process. Consequently, the target image has to be rotated counterclockwise $x°$ assuming that the target image has been classified under class x. After having rotated, the target image will be registered with the reference image.

4 Simulation Results

Pattern recognition network (*patternet*) which is part of MATLAB's Neural Network toolbox has been employed to learn and classify the synthetically generated reference image transformations. Patternet is essentially a feedforward that classify input as per the classes. The implementation of patternet demands that the class labels should consist of vectors of all zeros except for a 1 in location i where i is the class they are to represent. So, instead of 360 × 1 class label vector, we employ 360 × 360 matrix which is more of an implementation specific requirement.

Figure 1 shows a schematic of patternet architecture employed for the simulation experiments. As can be seen from the figure, 20 SURF features, form the input, with 30 hidden layers and 360 output classes accounting for 360 times 1° rotations.

The reference image, as detailed in the previous section, is rotated clockwise 360° and for each degree 20 strongest features were extracted using SURF feature extraction method [18]. So, the input data for patternet is essentially a 360 × 20 matrix and the supervisor signal for patternet is a 360 × 360 matrix containing class

Fig. 1 Patternet architecture employed in this study

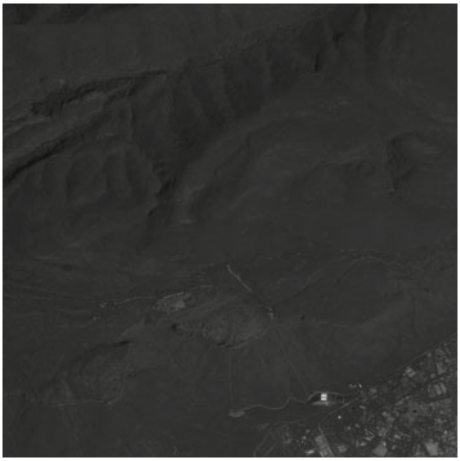


Fig. 2 Reference image

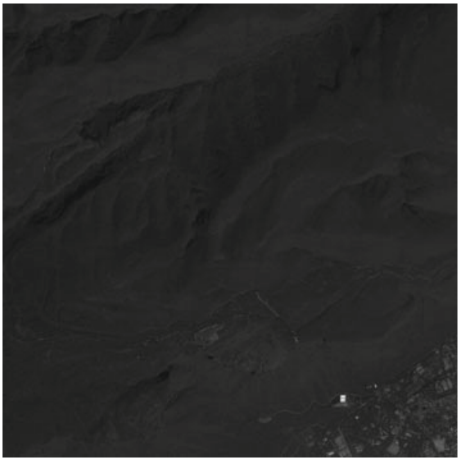


Fig. 3 Target image

labels for each of the 360 classes. The SURF features of target image (without any rotation) is also extracted to be used a target vector.

A pair of satellite images provided by National Remote Sensing Centre (NRSC), Hyderabad, India serves as the reference and target images. Figures 2 and 3 show the reference and target images used for the simulation experiments respectively.

Figure 4 shows the matched features between trained reference image and target image features. With the help of matched features, both image magnitude and direction of image is determined. We also experimented the rotation with 0.2° granularity. Table 1 shows the details of both the simulation experiments. It explained about 1° and 0.2° rotation. It took 350 iterations with gradient value of 1.84e-05 and performance value as 0.00749 with training time 57.54 min for 1° rotation. Similarly 389 iterations with gradient value of 6.42e-05 and performance value around 0.00748

Fig. 4 Image obtained from matching features at an angle of 90.20° rotation

Table 1 Simulation results

S. No.	Degree of rotation	No. of iteration	Gradient value	Performance value	Training time
1	1°	350	1.84e-05	0.00749	57:54 min
2	0.2°	389	6.42e-05	0.00778	11:26:52 h

Fig. 5 Registered image by 1° rotation

with training time 11:26:52 h for 0.2° rotation. Figures 5 and 6 respectively show the registered images after 1° and 0.2° rotation experiments. Finally Figs. 5 and 6 shows the registration of image at 1° and 0.2° rotation.

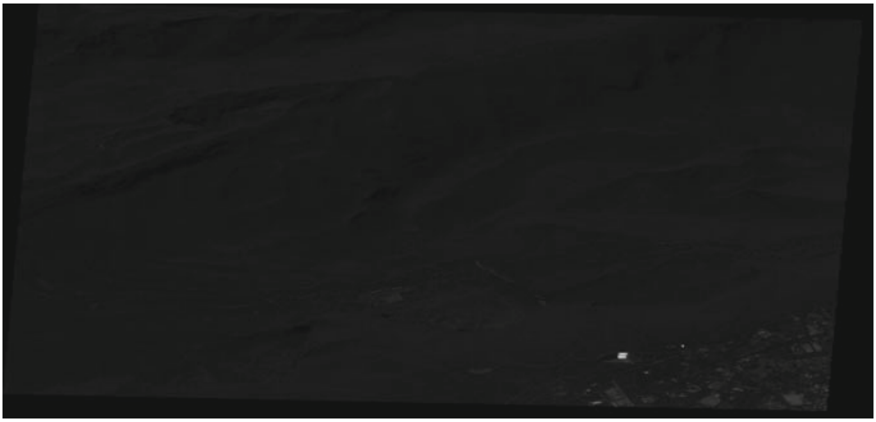


Fig. 6 Registered image by 0.2° rotation

5 Conclusion

In this paper, we have employed a simple and effective method for image registration using neural networks. The method involved generating and learning exhaustive synthetic image transformations (primarily rotation). Essentially, the reference image is rotated every 1° for 360 times and SURF features are extracted for each rotation with each rotation labelled as a class. Thus image registration is formulated as a classification problem. Once the patternet (the neural network employed) learnt the rotation classes, the target image, as is without any rotation, has been tested with the neural network to estimate the rotation the image might have undergone during the imaging process. After undoing the rotation (by rotating the image counter clockwise) the target image has been registered with the reference image. This registration method has been tested with satellite imagery. The granularity of the rotation as well as considering synthetic affine transformation will be part of the future work.

References

1. Menon HP, Narayanakutty KA (2010) Applicability of non-rigid medical image registration using moving least squares. Int J Comput Appl 1(6):79–86
2. Menon HP, Narayanakutty KA, Indulekha TS (2014) Feature point selection using structural graph matching for MLS based image registration. Int J Comput Appl Found Comput Sci 100(4):18–23
3. Balamurugan C (2013) Automated feature extraction from high resolution satellite imagery using ANN. Asian Rev Civil Eng 1:1–6
4. Gautier C, Somervile RCJ, Volfson LB (1986) Pattern Recognition of Satellite cloud imagery for improved weather prediction. National Aeronautics and Space Administration
5. Bhadran B, Nair JJ (2015) Classification of pattern on high resolution SAR images. In: Conference on computing and network communication. IEEE, pp 784–792

6. Wyawahare MV, Patil PM, Abhyankar HK (2009) Image registration techniques: an overview. Int J Signal Process 2:11–28
7. Xie Z, Pena JT, Gieseg M, Liachenko S, Dhamija S (2015) Segmentation by Surface to Image Registration. Virtual Scopics
8. Bulow H, Dooley L, Wermser D (2000) Application of principal axes for registration of NMR image sequence. Pattern Recognition Letters
9. Rao YR, Prathapani N, Nagabhooshanam E (2014) Application of normalized cross correlation to image registration. Int J Res Eng Technol 3:12–16
10. Kosinski W, Michalak P, Gut P (2012) Robust image registration based on mutual information measure. J Signal Inf Process 3:175–178
11. Mekky NE, Abou-Chadi FE-Z, Kishk S (2011) Wavelet based Image registration technique: a study of performance. Int J Comput Sci Netw Secur 11:188–196
12. Costin H, Rotariu C (2011) Medical Image Processing by using soft computing and information fusion. WSEAS international conference on mathematical methods and computational techniques and intelligent system
13. Wang S, Lei S, Chang F (2008) Image Registration based on neural network. In: International conference on information technology and application in biomedicine. IEEE, pp 74–77
14. Sarnel H, Senol Y, Sagirlibus D (2008) Accurate and robust image registration based on radial basis neural network. In: International symposium on computer and information science. IEEE
15. Abche AB, Yaacoub F, Maalouf A, Karam E (2006) Image registration based on neural network and fourier transform. In: Engineering in medicine and biology society. IEEE, pp 4803–4806
16. Krishna S, Varghese A (2015) Feature based automatic multiview image registration. Int J Comput Sci Softw Eng 4:308–314
17. Ariputhiran G, Gandhimathi S (2013) Feature extraction and classification of high resolution satellite images using GLCM and back propagation techniques. Int J Eng Comput Sci 2:525–528
18. Karami E, Prasad S, Shehata M (2015) Image matching using SIFT, SURF, BRIEF, and ORB: performance comparison for distorted image. In: Newfoundland electrical and computer engineering conference

Design and Development of Efficient Algorithms for Iris Recognition System for Different Unconstrained Environments

M. R. Prasad and T. C. Manjunath

Abstract One of the important concepts of identification of human beings in various sectors across the universe is the biometrics. In this paper, a brief report of the biometric recognition is being presented in a nutshell. This paper gives a brief conceptual view of the research work done on the topic titled, "Design and Development of Efficient Algorithms for Iris Recognition System for Different Unconstrained Environments" as the research topic chosen.

1 Introduction to the Research Work

One of the important concepts of identification of human beings in various sectors across the universe is the biometrics [1, 2]. This biometrics is defined as the art of identifying a human being by different methods. Identifying or verifying one's identity using biometrics is attracting considerable attention in this modern day automated world, one of the main reasons being the security issues in various places. Biometrics is the beautiful science of automatic identification of individuals that uses the unique physical or behavioral traits/characteristics of individuals to recognize them. Since biometrics is extremely difficult to forge and cannot be forgotten or stolen, biometric authentication offers a convenient, accurate, irreplaceable, and high secure alterna-

M. R. Prasad (✉)
Department of CSE, VTU RRC-Belagavi, Belgaum, Karnataka, India
e-mail: mrp.prasad@gmail.com

M. R. Prasad
Department of Computer Science and Engineering, JSS Academy of Technical Education (JSSATE), Dr. Vishnuvardan Road (Kengeri-Uttarahalli Road), Srinivapura Post, Bengaluru 560060, Karnataka, India

T. C. Manjunath
Department of ECE, Dayananda Sagar College of Engineering, Bangalore, Karnataka, India
e-mail: dr.manjunath.phd@ieee.org

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_89

tive for an individual, which makes it have more advantages over the traditional cryptography-based authentication schemes [3].

2 Brief Insight into the Literature Survey

In the recent digital era of the current central government, biometrics has been made compulsory in all the places (e.g., UID Aadhar, fingerprint, PAN, etc.). We had seen that even though there were lot of biometric methodologies, each one was suffering from one or the other drawbacks. Finally, in this context after studying the implications of each of the biometric methodologies, we arrived at the selection of the iris as the best method of biometric identification of human beings due to its large number of advantages [4]. Also, due to the current initiative taken up by the central government in the field of biometrics to be implemented in all sectors, this had further motivated us to take up the research work on the iris biometric field. This has made us identify the problem because of its vast application in each and every sector in this automated digital world [5].

Hence, in continuation, with zeal of this research work after making a thorough survey, we proposed some methodologies for the automatic recognition of biometric using iris by developing some algorithms in Matlab/LabVIEW, the problem finally, being defined as the research problem statement as "Design and Development of Efficient Algorithms for Iris Recognition System for Different Unconstrained Environments" with a brief analysis and the same was presented in the form of an exhaustive literature survey.

In majority of the work done by the various researchers in the relevant field [6–9], there were certain drawbacks, disadvantages, lacunas such as high computation to achieve good accuracy, longer execution time, large amount of storage to achieve secure authentication, and methodology failed when there was a noise like reflection in the image, thereby producing complex or extra edges of images; in fact, few works were done on increasing the performance and accuracy of the system w.r.t. high-speed computations. Some of the drawbacks were considered in our work and algorithms were developed, which was verified through effective simulation results in the Matlab and LabVIEW environments (software and hardware implementation) along with an application development incorporating the microcontroller for implementation.

3 Objective of the Research Work

The main objective of the research work is to develop image processing algorithms in Matlab/LabVIEW environments for the biometric identification of human beings through the iris part of the human eye under unconstrained environments [2, 10]. The other objective being to develop high-speed and efficient algorithms to overcome the security and recognition problems faced in many of the existing biometric authentication fields, to develop a system to work properly with all the types of human eyes and to improve the performance of the commonly used existing algorithms. The main motto of our research work is to make use of the iris images taken under unconstrained environments [11–14] and to develop algorithms to correctly identify an individual is present or not. The major scope of the proposed research work is mainly to develop some hybrid algorithms to overcome the security and recognition problems faced in many fields of the works as mentioned above, to develop an iris recognition system to provide fast identification and recognition for all human beings, to work properly with all eyes of different sizes using different types of classifiers, and finally to improve the performance of the commonly used existing algorithms.

4 Proposed Block Diagrams for the Recognition Process

The proposed overall block diagram of the developed authentication system makes uses of only two phases, viz., the iris enrolment/training phase and the iris recognition/evaluation/testing phase [15]. The iris images used in the research work were acquired from the CASIA-IrisV4 databases, which consists of all types of constrained and unconstrained image datasets. The complete iris scan recognition system could be summarized as consisting of different blocks with each block having its own functionality and all the blocks are used in our research work. They are the database (general/generated one), image acquisition/capturing, grayscale conversion, identification of ROI [16], preprocessing, resizing, boundary detection, segmentation [17–20], localization, normalization, noise removal, enhancement, feature processing, feature extraction, feature encoding, matching, classifiers, testing, decision taking, authentication, identification, recognition/matched, and the non-recognition/unmatched blocks.

The time consumption of the biometric authentication system developed was also very low, as it can identify an iris within few seconds. This compilation time includes all the times that were taken by the various processes mentioned precedingly. In fact, the proposed algorithms were worked out in the abovementioned phases. Successful simulations and experimentations were carried out, and encouraging results were achieved, thus claiming that the proposed systems are capable of fast and efficient iris identification over the existing ones under extreme conditions. In this context, it is iterated that nine contributions were developed as a part of the research work undertaken in the field of iris biometric recognition systems under unconstrained environments. Those nine contributions involved six in Matlab and three in LabVIEW, out of which one is a LabVIEW application of the voting process followed by the hardware implementation of the iris recognition system with a microcontroller using LabVIEW.

5 Proposed Identification Schemes in Seven Stages

The proposed identification scheme involved seven important stages in the detection process, viz.,

Stage 1 Preprocessing, segmentation, normalization, and enhancement,
Stage 2 Feature extraction,
Stage 3 Classification or matching,
Stage 4 Functional block diagram,
Stage 5 Overall working of the developed IRS,
Stage 6 Simulation results and the finally,
Stage 7 Conclusions.

The flow of our implemented research work is explained as follows.

In stage 1, the input is an eye image of a person who wants to get the authorization to the iris recognition system and the next step is preprocessing of the image considered (taken from CASIA-V4 database). In preprocessing step, the original image is enhanced and then the image iris is localized; once after localizing, the iris of the image pupil is localized by removing the occlusion present in the eye image. The various preprocessing operations are done with the help of Canny edge detection, Hough transforms, Fuzzy trapezoidal, Sobel operator method, OTSU algorithm, morphological operators, and the boundary detection methods. It has to be noted that in our work, the hybrid combination of preprocessing and segmentation concepts were used to obtain high degree of accuracy. In this content, normalization, segmentation, enhancement, and noise removal are also carried out, thus obtaining a noise-free, good resolution image that could be used further to extract its features [21, 22].

Coming to stage 2, which is the feature extraction stage, the various features of iris images are extracted using the concepts of local binary pattern features method, Gabor wavelets, log Gabor convolution wavelet method, hybrid SVD decomposition method, SFTA method, gray-level co-occurrence matrix method, fruit fly with cuckoo search algorithm, heuristic algorithm, local binary pattern features method, and the Haar wavelet methods. Once the features are extracted, it is stored in a refined database, sometimes being called the knowledge database. It has to be noted that in the work considered, the hybrid combination of different feature extraction methodologies was used to obtain high degree of accuracy. After all, the abovementioned steps in stage 2 were carried out to compare the feature vectors of the iris image with already stored feature vectors in database as templates (reference code).

So, the final step was to classify the test iris image by using different types of classifiers to classify the extracted iris into recognized ones and the unrecognized ones, which was considered in stage 3. Also in some cases, instead of classifiers, the matching techniques have been used to match the test iris with the already stored iris in the database. The different types of classifiers that were used for classification purposes are the SVM, ANNs, Multi-SVM, neural networks, and the radial basis functions. The different types of matching techniques that were used for matching purposes are the Hausdorff matching, surf matching, and the Hamming distance methods. Thus, by using these classifiers as a result if the stored feature vectors and the extracted template of test iris image are one, the same person is verified and authenticated, else it is rejected. It has to be noted that in the work considered, the hybrid combination of classifiers was used to obtain high degree of accuracy. If the match is found, then the iris is recognized, and the person is found in the database, else it is not recognized, i.e., the person is not there in the database.

In stage 4, all the block diagrams for the various contributions are proposed using hybrid combinations of the various processes used in the first three stages, followed by the brief working of the recognition process in stage 5. Codes are developed in the Matlab environment as .m files or .vi files in stage 6.

The developed .m/.vi files are run, the test query image is given as the input to the developed code, and after running the simulation, the simulation results are observed for the various contributions for the two cases, viz., matched (recognized) and unmatched (not recognized). Images which are captured under different unconstrained environments such as bad light, bad illumination, iris captured at a distance, at an angle or having parallax error, squint eyes, tilted iris, wearing glasses, eyes affected with diseases or having cataract done, etc. are given as the input to the proposed algorithms which are compared with the iris part which is already in the database; if the match occurs, then the person is verified or authenticated or matched or recognized, else it is rejected. Other type being the classifier approach used to classify the recognized irises and the unrecognized irises. All these methods proved to be more effective when the algorithms were developed and compared with the existing ones, thus establishing the supremacy of our methods over the others.

The proposed research work gives a simple and stable solution for iris recognition system for secure authentication in unconstrained environments. Although there are numerous works done on the chosen subject [23, 24], the proposed work stands unique with its ability of providing excellent results for various unconstrained cases of images considered in this work. The proposed algorithm/s focusses on the methods for rapid and accurate iris identification and authentication, especially in the field of preprocessing, feature extractions, and in the case of classification and matching of the irises. Satisfying and convincing results are obtained for all the defined objectives of the proposed algorithms which could be seen in a separate section of the simulation results presented under each contribution. The different algorithms which have been developed have provided very good results when it was implemented on CASIA-Iris Version 4 database along with better accuracy, high performance, and less error rate. At the end, also the results have been compared with some of the existing technologies and w.r.t. the work done by other authors to establish the supremacy of the proposed methods with them.

All the different proposed methods developed gave satisfying results for the recognition process, but the performance metrics, accuracy, differed among them. Performance metrics was also plotted to judge the best methodology. At the same time, in each contribution, a user-interactive automatic graphical user interface (GUI) was also developed as a part of the research work. This GUI takes the test pattern, compares with the database available, and if it is existing, immediately authenticates it well within couple of seconds, else it rejects the test pattern saying not authenticated or recognized, thus concluding the iris recognition process in the final stage presented as "conclusions" in stage 7.

6 Brief Insight into the Iris Image Databases

A brief review of the iris image databases that are being used in our research work was also explored. In fact, an image database can be defined as a collection of image data in .jpeg format, and typically all the images are associated with the activities of one or more related organizations, and thus the databases focus on the organization of images and its metadata in an efficient manner. There are different types of image databases and in our work, we have mainly concentrated on the CASIA-IrisV4 database for the analysis purposes.

Next, the implementation of the automatic biometric iris recognition system under unconstrained environments using the proposed methodologies is being presented in a nutshell. It also describes various steps that are used in the proposed methodology. In order to achieve the better accuracy, performance, and error rate than the existing methods, nine different iris recognition system techniques (software and hardware approaches) have been proposed which involve different preprocessing, feature extraction technique, and matching/classification algorithms. The entire work is presented in three stages in this research paper, viz.,

- design of the algorithms for iris recognition,
- the simulation results, and
- the development of an automated GUI for iris recognition.

7 Contributions of the Research Work

A brief introduction to the nine contributed works: The first six contributory works (software implementations in Matlab), next two contributory works (software implementations in LabVIEW) followed by the last contributory work (hardware implementations using a LabVIEW coupled with C), uses the concepts of preprocessing, edge detection, segmentation, normalization, feature extraction, matching, and classification for the recognition of an iris of any human being under the unconstrained environments.

Preprocessing (P), segmentation (S) and normalization (N) use the following:

1. PSN 1: CED—Canny Edge Detection and CHT—Circular Hough Transforms,
2. PSN 2: HT—Hough Transform and CED—Canny Edge Detection,
3. PSN 3: FTM—Fuzzy Trapezoidal and SOM—Sobel Operator Method,
4. PSN 4: OTSU—OTSU Threshold Values and BDM—Boundary Detection Method,
5. PSN 5: MO—Morphological Operators,

6. PSN 6: CED—Canny Edge Detection method and HT—Hough Transforms,
7. PSN 7: BD—Boundary Detection method,
8. PSN 8: BD—Boundary Detection method, and
9. PSN 9: MO—Morphological Operators.

Feature extraction (FE) use the following:

1. FE 1: LBP—Local Binary Pattern method,
2. FE 2: GW—Gabor Wavelets,
3. FE 3: GCW—1D Log Gabor Convolution Wavelet and HSVD—Hybrid SVD,
4. FE 4: SFTA—SFTA method,
5. FE 5: GLCM method and fruit fly with cuckoo search algorithm, heuristic algorithm,
6. FE 6: LBP—Local Binary Pattern method,
7. FE 7: 2D Gabor Wavelets and Haar Wavelets,
8. FE 8: 2D Gabor Wavelets and Haar Wavelets, and
9. FE 9: LBP—Local Binary Pattern method.

Classification (CN) use the following:

1. CN 1: m-SVM—Multi-SVM,
2. CN 2: HD—Hamming Distance,
3. CN 3: HM—Hausdorff Matching and SM—Surf Matching,
4. CN 4: KDB—Knowledge Data Base and ANN—Neural Network Algorithm,
5. CN 5: RBFNN, Neural Network, and SVM,
6. CN 6: HDM—Hamming Distance Method,
7. CN 7: ANN—Artificial Neural Network,
8. CN 8: ANN—Artificial Neural Network, and
9. CN 9: ANN—Artificial Neural Network.

Overall working of the six proposed methodologies in Matlab: The overall working of the developed methodology/flow diagram of the proposed method, i.e., the flow of research work, is as shown in Fig. 1 and is summarized as follows.

- Input is an eye image of a person who wants to get the authorization to the system and the next step is preprocessing of the inputted image considered.
- In preprocessing step, the original inputted image is enhanced and then the image iris is localized, and once after localizing, the iris of the image pupil is localized by removing the occlusion present in the eye image; different preprocessing techniques such as CED, CHT, HT, FTM, SOM, OTSU, BDM, and the MO methods are used for processing before the features are extracted [25].
- Then, the various features of an iris image are extracted using hybrid combination of LBP, GW, GCW, HSVD, SFTA, GLCM, FF, and CSA methods [26].
- After all, in the abovementioned steps, these feature vectors of an iris image are then made ready to compare with the already stored feature vectors in database as reference templates (reference code).
- So, the final step is to classify the test iris image by using hybrid combinations of classifiers that are m-SVM, HD, HM, SM, KDB, ANN, and the RBFNN methods.

- Thus, by using these classifiers as a result, if the stored feature vectors and the extracted template of test iris image are one and the same, then person is verified and recognized, else it is not recognized.

The general data flow diagram employed in this work for the nine contributory works is shown in Fig. 1.

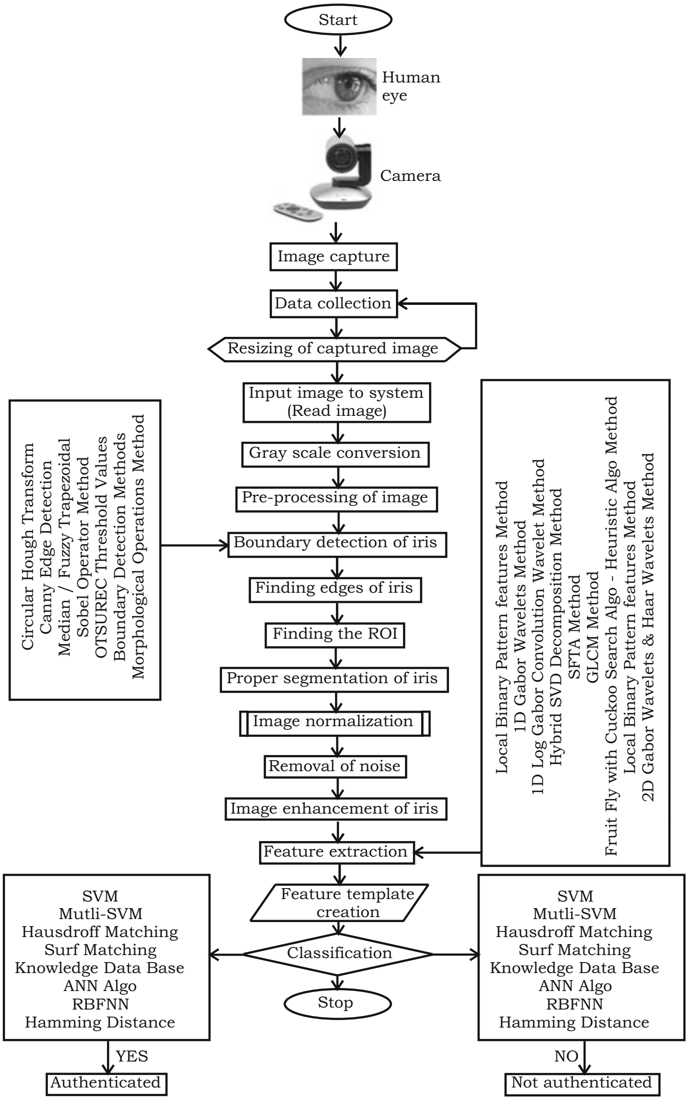


Fig. 1 General flowchart for the detection of iris in unconstrained environments for all the nine contributions

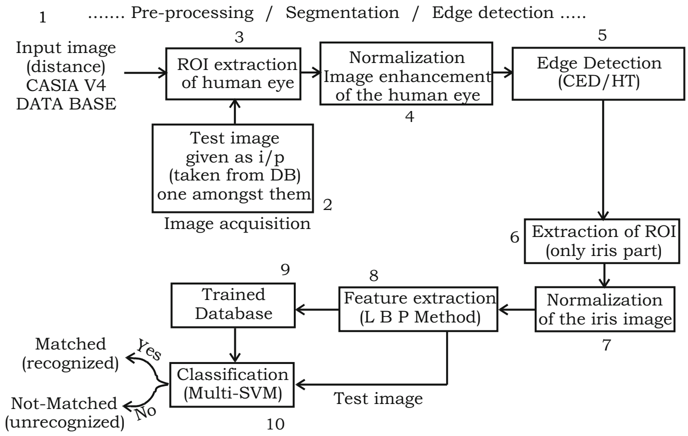


Fig. 2 Contribution 1—Iris recognition using CED/HT/LBP and Multi-SVM

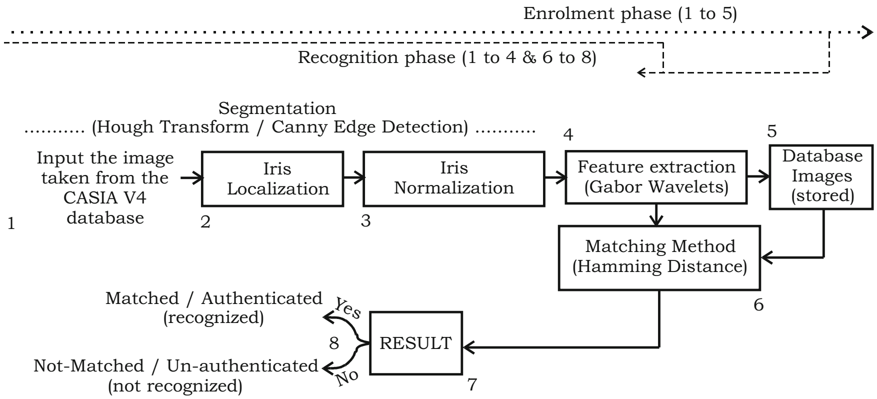


Fig. 3 Contribution 2—Iris recognition using (CED-HT)/(Gabor Wavelets)/(HD)

Contribution 1: Development of iris recognition using the concepts of preprocessing, segmentation (Canny Edge Detection/Circular Hough Transforms), feature extraction (Local Binary Pattern Features Method), and classification (Multi-SVM) of iris images with the development of an automatic GUI using Matlab. The proposed block diagram of this contributory work is shown in Fig. 2.
Contribution 2: Development of iris recognition using the concepts of preprocessing, segmentation (Hough Transform and Canny Edge Detection), feature extraction (Gabor Wavelets), and matching (Hamming Distance) of iris images with the development of an automatic GUI using Matlab. The proposed block diagram of this contributory work is shown in Fig. 3.
Contribution 3: Development of iris recognition using the concepts of preprocessing, segmentation (Fuzzy Trapezoidal and Sobel Operator Method), feature extrac-

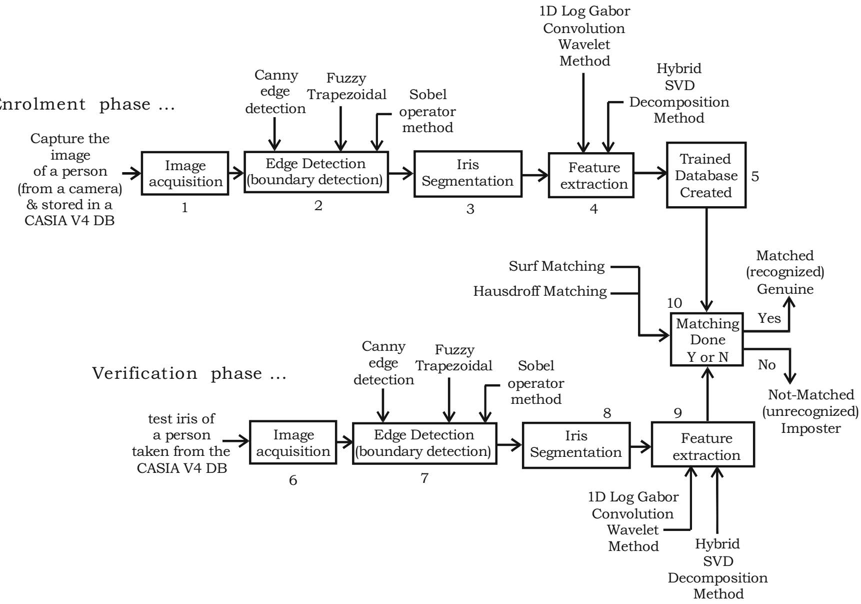


Fig. 4 Contribution 3—Iris recognition using (CED-Sobel-FTM)/(Gabor-SVD)/(Hausdorff–Surf method)

tion (1D Log Gabor Convolution Wavelet Method and Hybrid SVD Method), and Matching (Hausdorff Matching and Surf Matching) of iris images with the development of an automatic GUI using Matlab. The proposed block diagram of this contributory work is shown in Fig. 4.

Contribution 4: Development of iris recognition using the concepts of preprocessing, segmentation (OTSU Algorithm), feature extraction (SFTA Method), and Classification (Neural Network Algorithm) of iris images with the development of an automatic GUI using Matlab. The proposed block diagram of this contributory work is shown in Fig. 5.

Contribution 5: Development of iris recognition using the concepts of preprocessing, segmentation (Morphological Operators), feature extraction (GLCM Method and Fruit Fly with Cuckoo Search Algorithm, Heuristic Algorithm), and Classification (RBFNN, Neural Network, and SVM) of iris images with the development of an automatic GUI using Matlab. The proposed block diagram of this contributory work is shown in Fig. 6.

Contribution 6: Development of iris recognition using the concepts of preprocessing, segmentation (Canny Edge Detection Method and Hough Transforms), feature extraction (Local Binary Pattern Method), and matching of iris images (Hamming Distance Method) with the development of an automated GUI for iris biometric recognition using the Matlab tool [27–30]. The proposed block diagram of this contributory work is shown in Fig. 7.

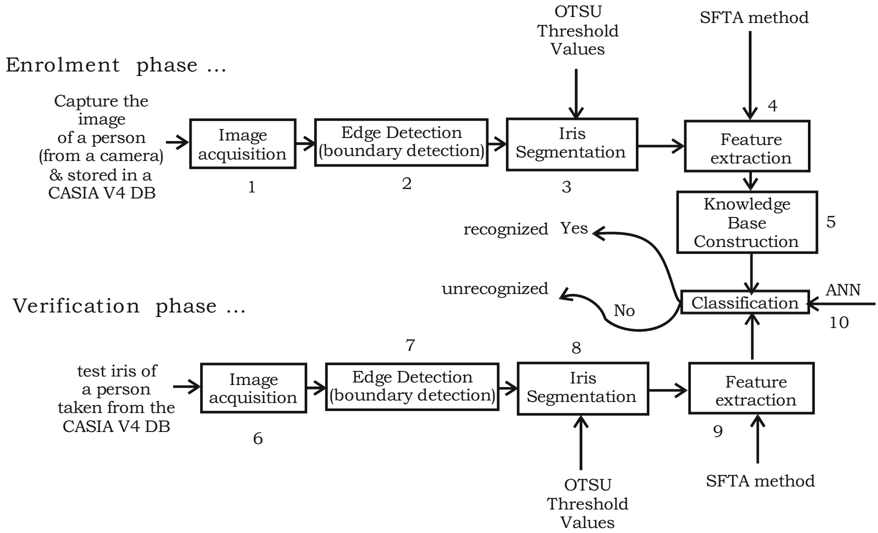


Fig. 5 Contribution 4—Iris recognition using (OTSU Algo)/(SFTA)/(ANN)

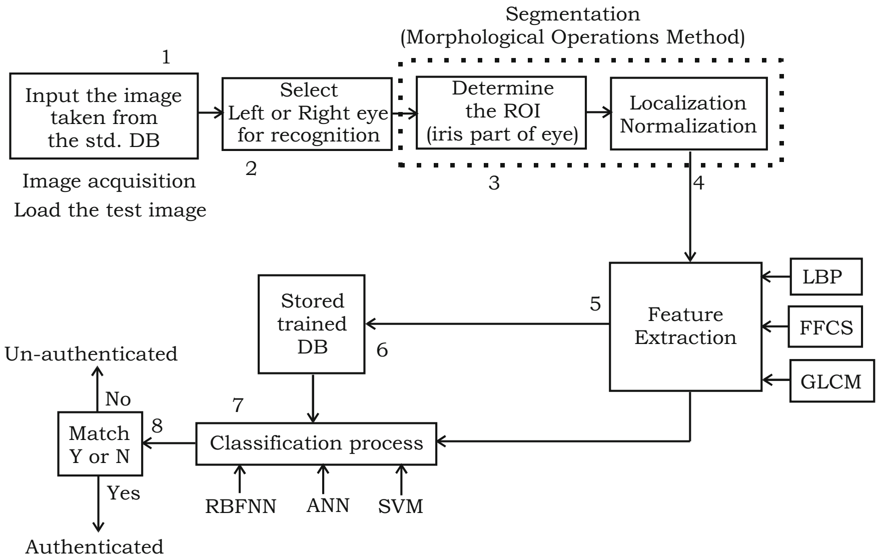


Fig. 6 Contribution 5—Iris recognition using (MO)/(GLCM-LBP-FFCS)/(RBFNN-ANN-SVM) method

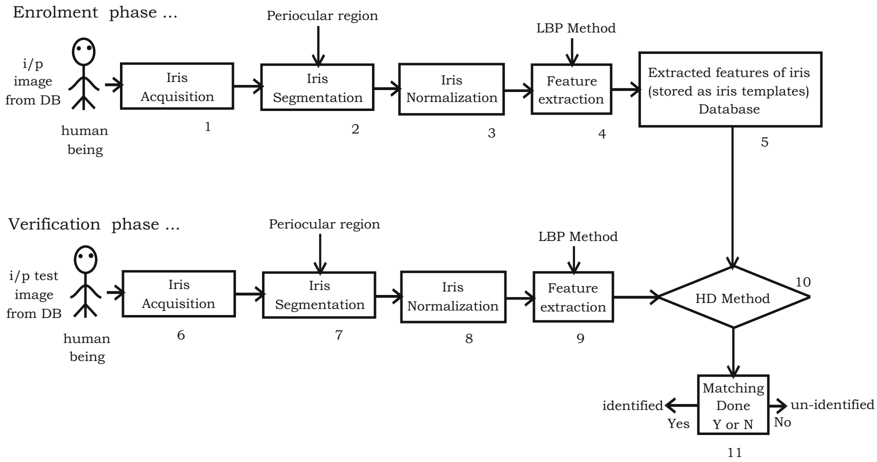


Fig. 7 Contribution 6—Iris recognition using (CED-HT)/LBP/HD method

8 Comparisons of the Contributions from C1 to C6

The six different methods which are proposed for the iris recognition are compared for their best performance, i.e., the comparisons of the different contributions from 1 to 6 are presented with their quantitative results for different cases of image processing such as segmentation, preprocessing, recognition, etc. It was found that for the entire set of images given from CASIA-V4 database, more than 99% of the images present in the database were recognized for all the six contributory works and the result of the percentage acceptance is shown in Table 1. Further, the accuracy of segmentation and recognition rate for all the contributions is shown in Table 1. Also, one of the contributions C5 has been compared with the work done by others which shows the supremacy of our method as shown in Table 2.

Table 1 Comparison rate of all the contributions

Method	Segmentation	Feature extraction	Matching/ classification	Percentage of segmentation	Percentage of recognition
1	CED/HT	LBP	SVM	99.8	99.72
2	CED	1D-GW	HDM	98.6	99.63
3	SO/Fuzzy/CED	1D-logGW/SVD	HDM/SURF	98.5	99.68
4	OTSU	SFTA	ANN	99.82	99.72
5	MO	GLCM/FFCS	RBFNN/ANN/SVM	99.7	99.85
6	CED/HT	LBP	HD	99.8	99.81

Table 2 Comparison rate of the proposed work with others

Algorithm	Correct recognition rate (%)
Daugman	99.5
Wildes	98
Boles	92.6
Li Ma	94.9
Roy	99.4
Ma	99.3
Liu	97.08
Avilla	97.89
Masek	83.92
Mayank	90
Proposed C5	99.85

9 Conclusion of the Six Proposed Methodologies (C1–C6)

The implementation results of the above six methodologies are, if the templates stored in the database is matched with the inputted image feature vector by using the hybrid methodology of classification as the classifiers, then it results into the conclusion that the person is recognized and genuine, else its imposter or not recognized (image not in the database).

10 Design and Developments Using the Contributions C7–C9

Contribution 7: Development of iris recognition using the concepts of preprocessing, segmentation (Boundary Detection Method), feature extraction (2D Gabor Wavelets and Haar Wavelets), and classification of iris images (ANN Method) with the development of an automated GUI for iris biometric recognition using the LabVIEW tool [31]. The proposed block diagram of this contributory work is shown in Fig. 8.
Contribution 8: Development of application of iris recognition system developed for electronic voting using LabVIEW. The proposed block diagram of this contributory work is shown in Fig. 9.
Contribution 9: Hardware implementation of the iris recognition concept using ATMEL microcontroller interfaced with LabVIEW with the development of an automated GUI. The proposed block diagram of this contributory work is shown in Fig. 10.

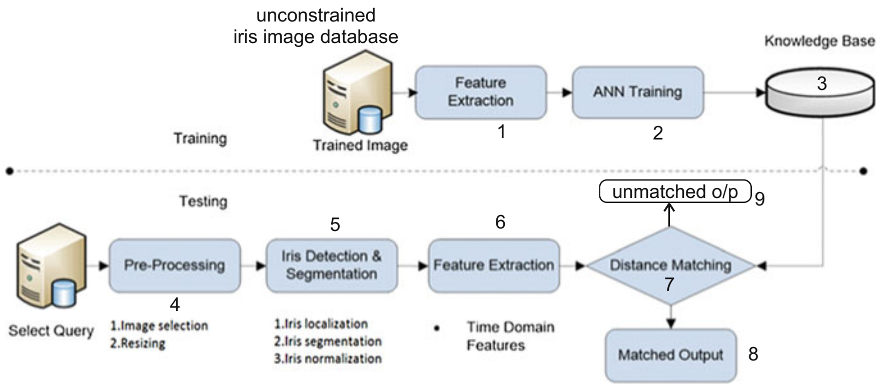


Fig. 8 Contribution 7—Proposed iris recognition BD using ANN with LabVIEW tool

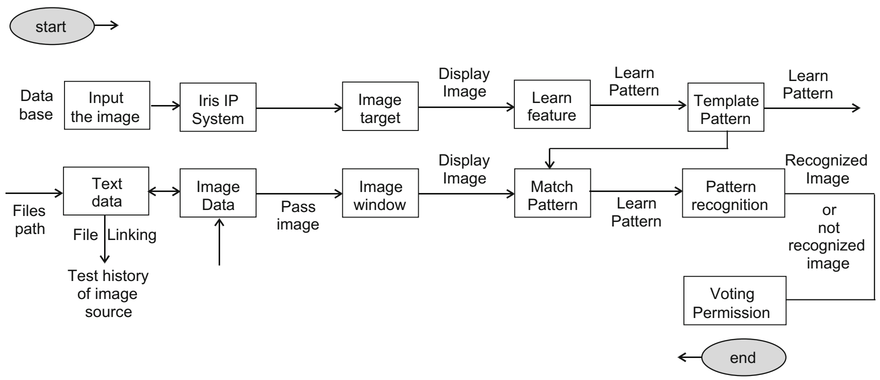


Fig. 9 Contribution 8—Data flow diagram of the IRS used for voting purposes

11 Working Principles of Contributions C7–C9

Working principle of C7: The seventh contributory work consists of developing an interactive iris recognition system for unconstrained images using the concept of preprocessing, segmentation (Boundary Detection Method), feature extraction (2D Gabor Wavelets and Haar Wavelets), and classification of iris images (ANN) with the development of an automated GUI for iris biometric recognition using the LabVIEW tool, and the iris recognition system is developed as per the block diagram shown in Fig. 8. Here, all the concepts of the various biomedical image processing such as the database generation, image acquisition/capturing, grayscale conversion, identification of ROI, preprocessing, resizing, boundary detection, segmentation, localization, normalization, noise removal, enhancement, feature processing, feature extraction, feature encoding, matching, classifiers, testing, decision taking, authentication, identification, and recognition are being used.

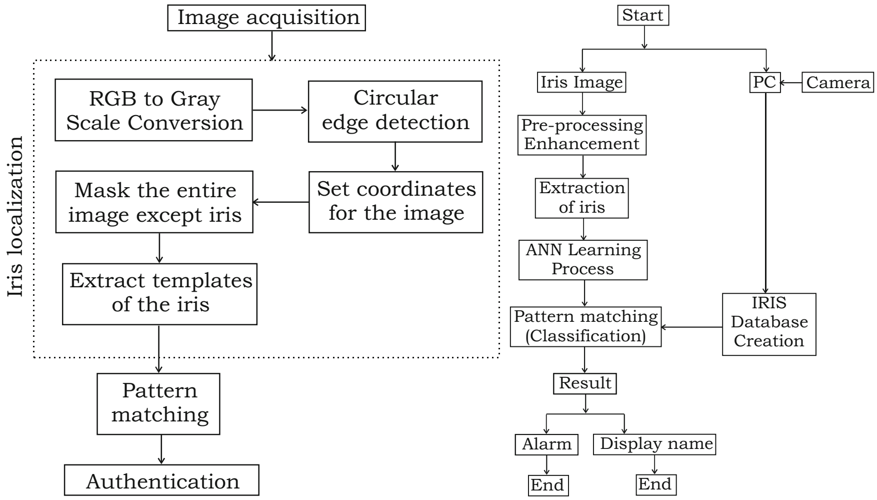


Fig. 10 Contribution 9—Algorithm for the hardware implementation using a μC

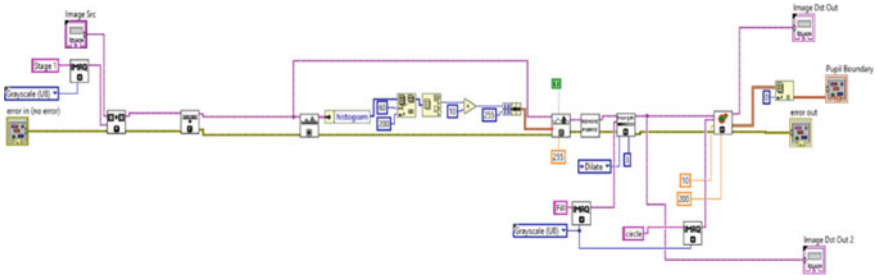
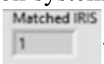
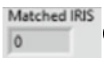


Fig. 11 Block diagram of the determination of pupil in LabVIEW (C7)

Circuit diagrams are being developed in the LabVIEW environment for each of the separate IP processes (such as pupil detection, limbic boundary detection, normalization, segmentation, resizing, feature extraction, noise removal, recognition, etc.) and then finally compiled into a hybrid model. One such diagram developed is shown in Fig. 11. When the overall hybrid model is run, the various simulation results are observed as shown in Figs. 12 and 13, respectively, thus showing the efficacy of the developed iris recognition system. If the test iris is matched with the iris in the database, then it appears as 

, else it appears as 

(Fig. 14).

Working principle of C8: Eighth contributory work is an application-oriented one, which includes the development of an intuitive iris acknowledgment framework for unconstrained pictures utilizing the idea of IP and implementing it for a voting application. Circuit diagrams are being developed in the LabVIEW environment for each of the separate IP processes (such as pupil detection, limbic boundary detection,

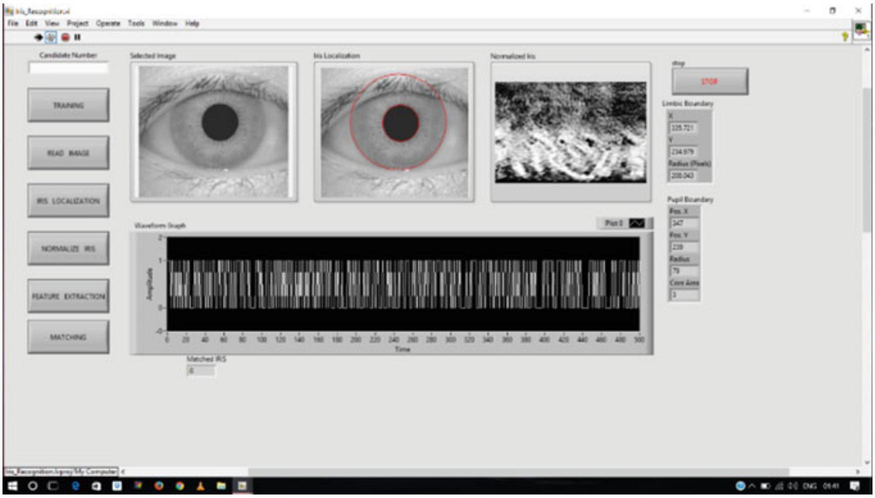


Fig. 12 GUI of the IRS application generated using LabVIEW—Matched iris (recognized—display shows 1)…(C7)

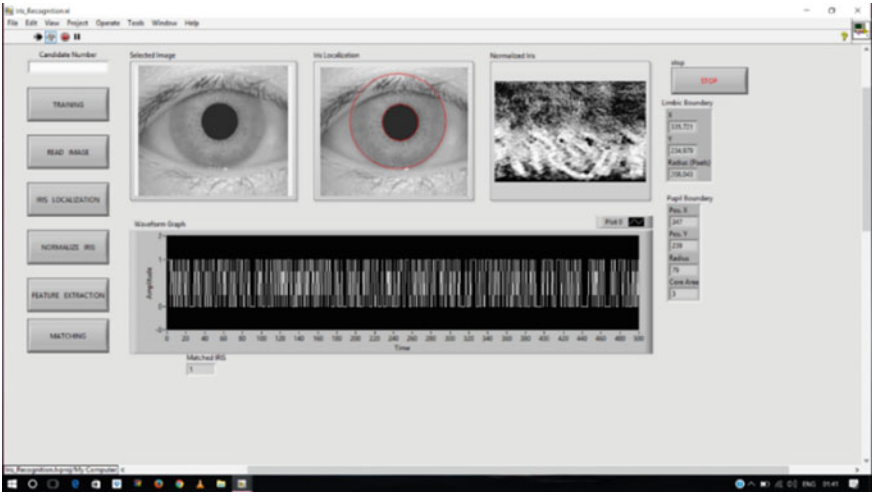


Fig. 13 GUI of the IRS application generated using LabVIEW—Not matched iris (not recognized—display shows 0)…. (C7)

normalization, segmentation, resizing, feature extraction, noise removal, recognition, etc.) for the considered voting application scenario and finally merged to form a hybrid model. To start with, all the iris images are taken from the standard available database one by one, preprocessing done, and iris parameters are being found out and saved in a DB. Next phase is called as verification or testing phase; the test iris which is to be detected to find whether its presence is there in the refined database or not is

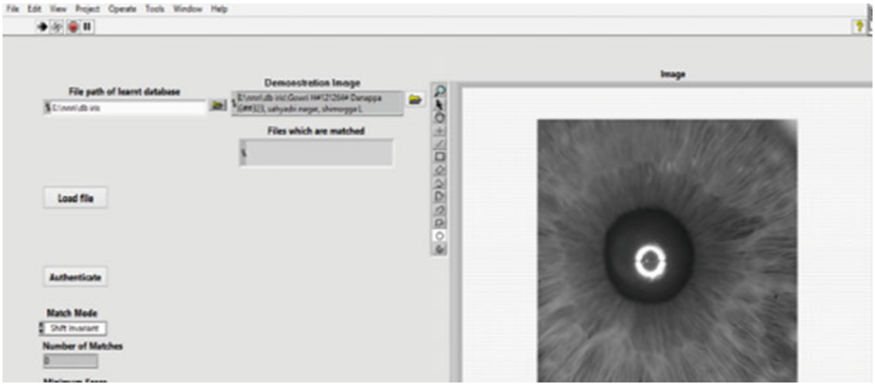


Fig. 14 Loading of the image of the human being (C8)

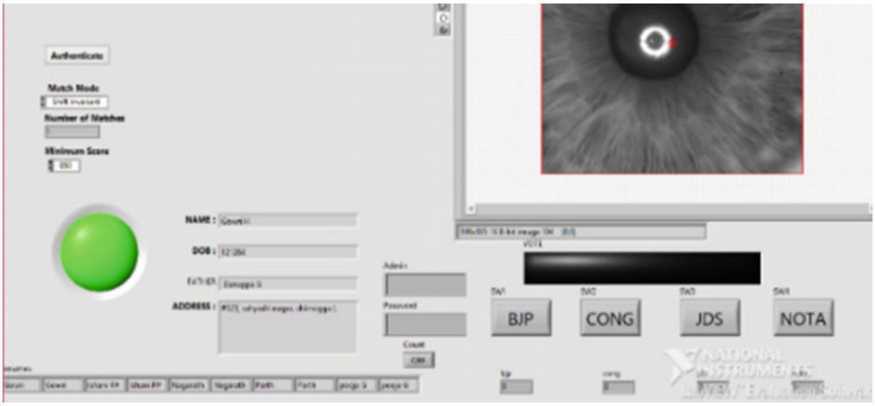


Fig. 15 Recognition success by an authorized person in the database (C8)

given as input to the proposed algorithm and all the processes as followed previously are carried out, and the matching of the test iris with its counterpart in the refined database is found out whether it is there or not using a user-friendly developed GUI.

If the recognition is successful, then the iris is authenticated (tab appears green in color), else it is not recognized (tab appears red in color). Now, once the iris is being recognized (Fig. 15), the person can do the voting by pressing the corresponding button against the party name, after which the red LED glows indicating that the person has voted (Fig. 16). For an unrecognized case, the test iris which is not there in the database is being loaded in the GUI and immediately the matching score says "0" indicating that the person is not present in the database and the tag glows red showing that the person is not there in the database and hence he cannot vote, proving that recognition is unsuccessful (Fig. 17), indicating red button.

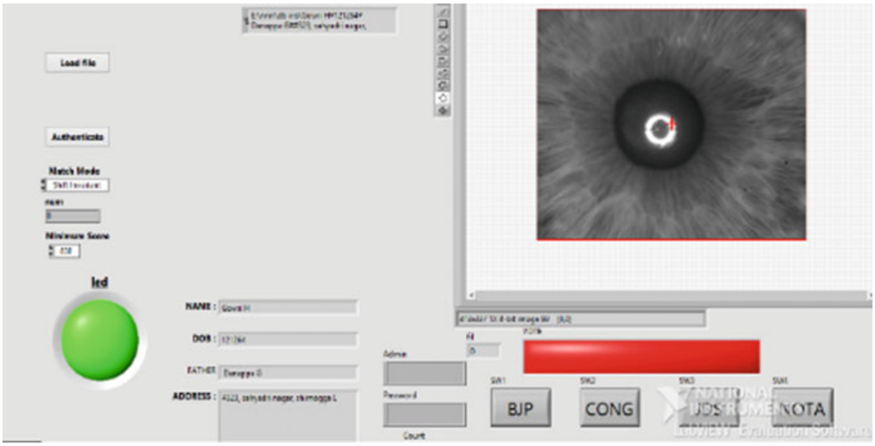


Fig. 16 Simulation results after being voted by a recognized human being (C8), shown in green color

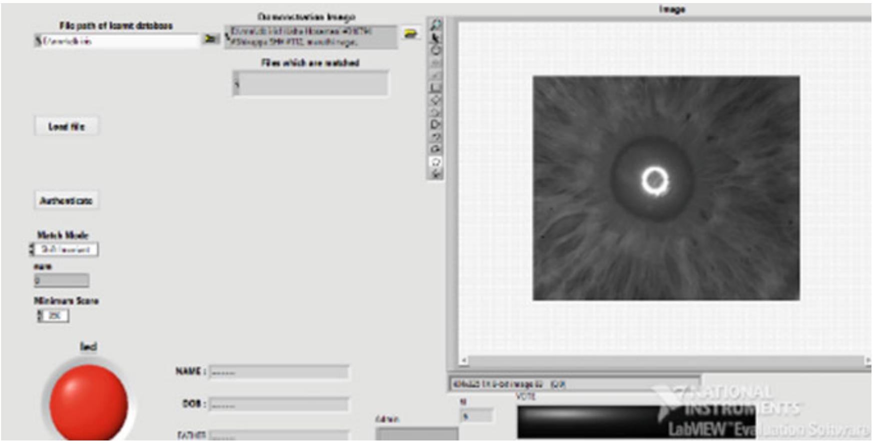


Fig. 17 Recognition unsuccessful by an unauthorized person in the database (failure)….(C8), shown in red color

Working principle of C9: The ninth contributory work involves the hardware implementation of the iris recognition process using an Atmel microcontroller with the simulation being done in the LabVIEW environment and then dumping the code onto the μC. All the concepts of the various biomedical image processing such as the database generation, image acquisition/capturing, grayscale conversion, identification of ROI, preprocessing, resizing, boundary detection, segmentation, localization, normalization, noise removal, enhancement, feature processing, feature extraction, feature encoding, matching, classifiers, testing, decision taking, authentication, identification, and recognition are being used [32–34]. Once the code has been developed,

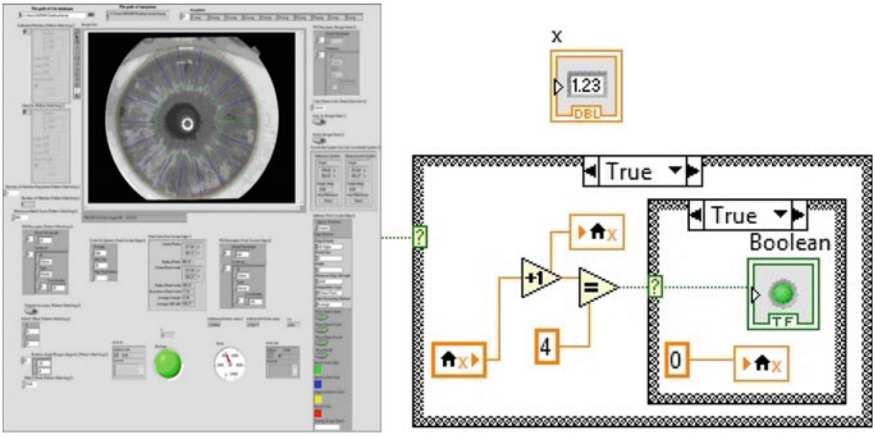


Fig. 18 VI G-code for recognition with recognized status shown in green color (C9) with the developed LabVIEW circuit diagram

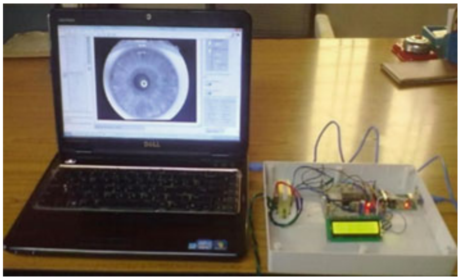


Fig. 19 Hardware interfaced to µC using a laptop equipped with an NI LabVIEW tool

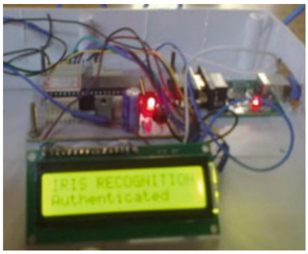


Fig. 20 Iris being recognized (authenticated) using the µC hardware kit (Contribution—9)

it is run and the various simulation results are observed as shown in Figs. 18, 19, and 20, respectively. It has to be noted that the iris images are being obtained from the standard iris image database. The LabVIEW tool is being interfaced to the µC, and the hardware results are observed on the display which is interfaced to the µC (Fig. 14).

Finally, the simulation results are validated with the hardware results justifying the same. The hardware unit of the IRS consists of the ATmega32 a Hi-Pi, LP AVR®

8-bit μC which is interfaced with a (16 × 2) Hitachi JHD 162A LCD display for displaying the iris recognition results. Also, a USB to UART converter adapter is being used for interfacing this hardware with the computer system using an SFE USB-RS232 Controller Driver. Along with this hardware, a 6-0-6 transformer, a 7805 5 V IC regulator, ST-232 driver circuits, USB cable, and power cords are being used for the hardware implementation purposes. The LCD display will display the authentication results obtained from the PC. The MAX-232 is a coordinated circuit, which gives the signal from an RS-232 serial port to the concerned signals for the use in a TTL good advanced rationale circuitry.

The MAX232 is a double drive/beneficiary and normally changes over the R_X, T_X, CTS, and the RTS signals. Figures 19 and 20 show the interfacing connections required to interface the ATmega32 with an LCD. The port A is used to get the o/p from microcontroller such that to display the test results in the display. The figure of hardware implementation (Figs. 19 and 20) is interfaced to a laptop/PC, which is fully loaded with NI LabVIEW s/w along with flowchart of the recognition process in Fig. 10. To start with, all the iris images are taken from the standard available database one by one, preprocessing done, extract iris features, train the different irises using ANN, and stored in a refined database. In the second phase, called as the verification or the testing phase, the test iris which is to be detected to find whether its presence is there in the refined database or not is given as input to the proposed algorithm and all the processes as followed previously are carried out and then, the matching of the test iris with its counterpart in the refined database is found out whether it is there or not. If the recognition is successful, then the iris is authenticated (LED turns green), else it is not recognized (LED turns red).

This work portrays a substitute strategy to distinguish people utilizing pictures of their iris using simulations as well as using hardware. Pattern recognition or template matching is carried out using the National Instruments LabVIEW tool. The developed NI LV script is finally run with the pictures which are inputted from the standard database. The GUI is developed using the NI LabVIEW and NI VISION Assistant is used for the extraction of templates characteristics. This recognizable proof framework is basic requiring just couple of parts and is sufficiently compelling to be coordinated inside security frameworks that require a personality check. Once simulations could be observed, then *G*-code is dumped on to the μC through the relevant ports and the recognition phenomenon is observed on the display.

12 Conclusions of Contributions C7–C9

Research was done in the area of iris recognition systems with the help of LabVIEW tool. Three different contributions (C7–C9) were carried out, with each one having one or the other advantages over the others. First one being a typical iris recognition system design using LabVIEW, second one being an application of iris recognition system using LabVIEW for a voting application, and the third one being the hardware implementation of IRS using LabVIEW. Simulations presented under the simulation results section in each contribution show the effectiveness of the methods proposed for developing a user-friendly graphical user interface for the iris recognition.

13 Overall Conclusions of the Research Work

In this section, the brief outcome, i.e., the conclusions about the research work done, followed by the scope for the future work in this exciting field of biometric authentication and identification under *unconstrained environments* is being presented.

Research was done for the efficient approach for secure biometric authentication using iris. Broad writing literature overview was being done in this emerging–developing field. In this context, a sincere attempt was being made to develop a simple and efficient method for iris recognition using simple segmentation method. The time consumption of the system is also very low, as it can identify an iris within few seconds. This time includes segmentation, feature extraction, feature selection/dimension reduction, and classification time. After the successful experiments and the encouraging results achieved, it can be claimed that the proposed system is capable of fast and efficient iris identification.

The proposed methodology developed by us has a wide range of advantages such as simple algorithm and easy to implement, less complicated, less execution time, faster authentication and recognition, iris has a fine texture because of which the extraction of features becomes easy, speed of the recognition system increases significantly, false acceptance ratio reduces up to a significant extent, performs robustly with different image qualities, highly reliable system for authentication, highly stable over lifetime, and many more. The recognition system developed is quite simple requiring only few components with the concept of the usage of hybrid methodologies and is highly effective enough to be integrated within security systems that require an identity check in any engineering application.

Research was done for the efficient approach for secure biometric recognition using iris under a large number of unconstrained environments such as tilted iris, iris captured under bad light illumination, from a distance, at an angle, wearing spectacles, occluded, etc.

In the proposed system, CASIA-V4 databases have been used for the recognition of iris in human beings and the work can be extended to other databases also in the near future by the future researchers. Moreover, average time consumption of the system could be improved by changing/improving the segmentation technique and other classifiers may also be used to evaluate the system. A critical feature of this coding approach is the achievement of commensurability among iris codes, by mapping all irises into a representation having universal format and constant length, regardless of the apparent amount of iris detail.

Any iris extraction from the acquired image needs four important parameters, viz., preprocessing, segmentation, feature extraction, and classifiers approaches. Different algorithms were used in the above four cases, each of which is mentioned below in greater detail. The various preprocessing methodologies used in our work are Hough

Transforms, Mean Median Std. Deviation and Entropy Methods, Morphological Operations Method, and the Boundary Detection Method.

The various segmentation methods that were used in our work for segmenting the iris image from the background are canny edge detection, fuzzy trapezoidal, and the SFTA—Segmentation-based Fractal Texture Analysis method, rising and falling edge method. The various feature extractions of iris images are done by using hybrid combinations of Gray-Level Co-occurrence Matrix (GLCM), LBP, FT, 2D-Gabor filter, and Fast Fourier Transform (FFT).

The various classifiers/matching concepts that were used in our work for classifying the recognized images are SVM, ANN, RBFNN, HD, and SURF which give effective results with less number of false detections (fake detection and recognition). Satisfying and convincing results are obtained for all the defined objectives of the proposed algorithms. Thus, biometric recognition system using iris which is developed by us is quite simple and requires few components and is effective enough to be integrated within security systems that require an identity check. The work can be used in security- and authentication-related applications for unconstrained environments successfully, and the errors that have occurred can be easily overcome by the use of stable equipment.

A. *Nine contributory works were carried out in this research work which was briefly explained in different sections.*
B. *The eight different algorithms provided good results when implemented on the iris CASIA Version 4 database under unconstrained environments.*
C. *One application problem was considered using LabVIEW and was successfully tackled upon with.*
D. *A hardware implementation of the iris recognition using a microcontroller was also implemented which yielded good results of iris authentication.*

Related to the methods mentioned in the exhaustive literature review, the proposed methods vary from different aspects and take into account all possible outcomes such as variations in the iris part in varying sunlight and brightness (*low light, high glares, darker regions, wearing spectacles, half eyes closed, half eye open, squint eyes, partially occluded eyes,* etc.). Positions of iris at different distances and angles [35–37], when iris part is not captured completely due to blink of eyelashes, presence of sunglasses, and other factors based on the work done to simulate and capture the iris image with different cases like angle, light incident on surface, and different other real-time scenarios are also considered.

The outcome of the research work is to show that when the designed algorithm/s developed in the Matlab/LabVIEW environment is run, the automatic recognition of the iris is done with minimum computational time in comparison with the work done by the other researchers till date taking into consideration many of the drawbacks of the fellow researchers, thus enhancing and improving the performance of the existing algorithms under unconstrained environments with the development of an automated GUI for iris recognition purposes. The nine different algorithms which were proposed provided good results when implemented on the different standard CASIS Version 4 iris databases under unconstrained environments.

14 Scope for the Future Work

At the end, the scope for future work, i.e., extensions of the research work that is being done, is being presented in this paper, which can be explored upon by the future research scholars. Iris recognition is an important and challenging task in secure authentication. Even though the proposed work gives good results with many advantages, many improvements can be done. Future work can aim to enhance the system to deal with image degradation by noise, duty, and glasses. Color feature can be also used to increase the recognition accuracy. There is scope to work on developing an algorithm based on multimodal biometrics system which will combine both the biometric characteristics derived from one or modalities such as palm print and iris which give high level of security. To conclude, a sincere attempt was being made to develop simple and efficient methods for iris recognition in "***unconstrained environments***" using a combination of several methodologies (hybrid algorithms) in the various processes of the iris biometric recognition which could be seen from the results of simulation and experiments.

References

1. Daugman J (2004) How iris recognition works. IEEE Trans Circuits Syst Video Technol 14(I):21–30
2. Daugman J (2004) Recognizing persons by their iris patterns. In: Proceedings of Advances in Biometric Person Authentication, vol 3338, pp 5–25
3. Daugman J (2007) New methods in iris recognition. IEEE Trans Syst Man Cybern Part B-Cybern 37(5):1167–1175
4. Daugman J (2003) The importance of being random: statistical principles of iris recognition. Pattern Recognit 36:279–291
5. Daugman JG, Downing CJ (2001) Epigenetic randomness, complexity & singularity of human iris patterns. Proc R Soc Lond B Biol Sci 268:1737–1740
6. Zhou Z, Du YZ et al (2009) Transforming traditional iris recognition systems to work in non-ideal situations. IEEE Trans Ind Electr 56(8):3203–3213
7. Daniel DM, Monica B (2010) Person authentication technique using human iris recognition.In: IEEE Conference Paper, pp 265–268. 978-1-4244-8460-7/10/$26.00©2010
8. Li P, Ma H (2012) Iris recognition in non-ideal imaging conditions. Pattern Recogn Lett 33:1–9
9. Karakaya M (2016) A study of how gaze angle affects the performance of iris recognition. Elsevier Sci Direct's Pattern Recog Lett 82:132–143
10. Fancourt C, Bogoni L, Hanna K, Guo Y, Wildes R, Takahashi N, Jain U (2005) Iris recognition at a distance. In: Proceeding of international conference on audio and video based biometric person authentication, vol 3546, pp 1–13
11. Moi SH, Asmuni H, Hassan R, Othman RM (2014) A unified approach for unconstrained off-angle iris recognition. In: International symposium on biometrics & security technologies (ISBAST), pp 39–44. 978-1-4799-6444-4/14/$31.00©2014. IEEE
12. Kaur N, Juneja M (2014) A novel approach for iris recognition in unconstrained environment. J Emerg Technol Web Intell Acad Publishers 6(2):243–246
13. Tsai Y-H (2014) A weighted approach to unconstrained iris recognition. World Acad Sci Eng Technol Int J Comput Inf Eng 8(1):30–33. ISSN:1307-6892
14. Shin KY, Nama GP, Jeong DS, Cho DH, Kang BJ, Park KR, Kim J (2012) New iris recognition method for noisy iris images. Elsevier Sci Direct's Pattern Recogn Lett 33:991–999

15. Chen Y (2010) A high efficient biometrics approach for unconstrained iris segmentation and recognition. Ph.D. Thesis. College of Engineering and Computing, Florida International University, Greater Miami
16. Roy K, Bhattacharya P et al (2010) Unideal iris segmentation using region-based active contour model. In: Campilho A, Kamel M (eds) ICIAR 2010, Part II. LNCS 6112, © Springer, Berlin, Heidelberg, Germany, pp 256–265
17. Jan F (2017) Segmentation and localization schemes for non-ideal iris biometric systems. Elsivier's Sci Direct Jr Signal Proc 133:192–212
18. Reddy N, Rattani A, Derakhshani R (2016) A robust scheme for iris segmentation in mobile environment. In: IEEE symposium on technologies for homeland security (HST). IEEE, Florida, USA. 978-1-5090-0770-7/16/$31.00©2016
19. Santoso A, Choirunnisa S, Prihasto B, Wang J-C (2016) Improving iris image segmentation in unconstrained environments using NMF-based approach. In: 2016 IEEE international conference on consumer electronics-Taiwan (ICCE-TW). IEEE, pp 27–29. 978-1-5090-2073-7/16/$31.00©2016
20. Yahiaoui M, Monfrini E, Dorizzi B (2016) Markov Chains for unsupervised segmentation of degraded NIR iris images for person recognition. Elsevier Sci Direct's Jr Pattern Recogn Lett 82:116–123. http://dx.doi.org/10.1016/j.patrec.2016.05.025
21. Sahmoud SA, Abuhaiba IS (2013) Efficient iris segmentation ethod in unconstrained environments. Elsivier's Science Direct Jr. of Pattern Recognition, vol 46, pp 3174–3185
22. Mahlouji1 M, Noruzi A (2012) Human iris segmentation for iris recognition in unconstrained environments. IJCSI Int J Comput Sci Issues 9(1), No. 3:149–155. ISSN (Online): 1694-0814
23. Proenca H (2010) An iris recognition approach through structural pattern analysis methods. Expert Syst 27(1):6–16
24. Raffei AFM, Hishammuddin A, Hassan R, Othman RM (2013) Feature extraction for different distances of visible reflection iris using multiscale sparse representation of local Radon transforms. Elsivier's Sci Direct Jr Patt Recogn 46:2622–2633
25. Haindl M, Krupička M (2015) Unsupervised detection of non-iris occlusions. Elsevier Sci Direct's Pattern Recogn Lett 57:60–65
26. Barpanda SS, Majhi B, PankajKumar S (2015) Region based feature extraction from non-cooperative iris images using triplet half-band filter bank. Elsevier Sci Direct's Jr Opt Laser Technol 72:6–14
27. Tan C-W et al (2013) Towards online iris & periocular recognition under relaxed imaging constraints. IEEE Trans Image Proc 22(10):3751–3765
28. Santos G, Hoyle E (2012) A fusion approach to unconstrained iris recognition. Elsevier Sci Direct's Pattern Recogn Lett 33:984–990
29. Proença H, Neves JC (2016) Visible-wave length iris/periocular imaging and recognition surveillance environments. Elsevier Sci Direct's Jr Image Vis Comput 55:22–25
30. Liu B, Lam S-K, Srikanthan T, Yuan W (2011) Utilizing dark features for iris recognition in less constrained environments. In: Fourth IEEE international symposium on parallel architectures, algorithms and programming. IEEE Comp Soc, Tianjin, China, pp 110-114. ISBN 978-0-7695-4575-2/11 $26.00©2011, https://doi.org/10.1109/paap.2011.51
31. Hajaria K, Gawandeb U, Golharc Y (2016) Neural network approach to iris recognition in noisy environment. Elsivier's Sci Direct Int Conf Inf Secur Priv (ICISP2015) Procedia Comput Sci 78:675–682, 11–12. https://doi.org/10.1016/j.procs.2016.02.116.Nagpur, India
32. Liu J, Sun Z, Tan T (2014) Distance metric learning for recognizing low-resolution iris images. Elsevier Sci Direct's Jr Neurocomp 144:484–492
33. Alvarez-Betancourt Y, Garcia-Silvente M (2016) A key points—based feature extraction method for iris recognition under variable image quality conditions. Elsevier Sci Direct's Jr Knowl-Based Syst 92:169–182
34. Hu Y, Sirlantzis K, Howells G (2015) Exploiting stable and discriminative iris weight map for iris recognition under less constrained environment. In: 7th IEEE international conference on biometrics theory, applications and systems (BTAS), Arlington, VA, USA, ISBN: INSPEC Accession Number: 15668109, https://doi.org/10.1109/btas.2015.7358759, 8–11 Sept 2015

35. Hu Y, Sirlantzis K, Howells G (2017) A novel iris weight map method for less constrained iris recognition based on bit stability and discriminability. Elsevier Sci Direct's Jr Image Vis Comput 58:168–180
36. Arya KV, Gupta A, Kumar G, Singhal P (2012) A novel approach to minimize the impact of non ideal samples in iris recognition system. In: Third IEEE international conference on computer and communication technology. IEEE Comp Soc, Allahabad, U.P., India, pp 352–356. 978-0-7695-4872-2/12 $26.00©2012 IEEE, https://doi.org/10.1109/iccct.2012.77, 23–25 Nov 2012
37. Perez C, Lazcano V, Estévez P, Held C (2009) Real-time template based face and iris detection on rotated faces. Int J Opto-mech 3:54–67. Taylor & Francis Group, UK. https://doi.org/10.1080/15599610902717801, ISSN: 1559-9612 (Print) 1559-9620 (Online)

Relation Extraction Using Convolutional Neural Networks

V. Hariharan, M. Anand Kumar and K. P. Soman

Abstract Identifying the relationship between the entities plays a key role in understanding any natural language. The relation extraction is a task, which finds the relationship between entities in a sentence. The relation extraction and named entity recognition are the subtasks of information extraction. In this paper, we have experimented and analyzed the closed-domain relation extraction using three variants of temporal convolutional neural network on SemEval-2018 and SemEval-2010 relation extraction corpus. In this approach, the word-level features are formed from the distributed representation of text and the position information of entity are used as the feature for the model.

1 Introduction

Ever since the Internet outbreak, there has been a tremendous rise in the amount of information that has been found on the various online platforms like Reddit, Stack Exchange, etc. To study and understand this information on a large scale in various platforms, it is necessary to build an information extraction system for qualitative data analysis. The information extraction is a task of extracting the factual information from the unstructured text. This task can be done jointly or can be divided into two subtasks such as Named Entity Recognition (NER) and relation extraction. In the NER task, named entities in the text are identified and each entity is assigned to a single category [1]. After recognizing the entities, finding the semantic relation between them is called relation extraction. This work focuses toward solv-

V. Hariharan (✉) · M. Anand Kumar (✉) · K. P. Soman (✉)
Center for Computational Engineering and Networking (CEN), Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Coimbatore, India
e-mail: cb.en.p2cen16007@cb.students.amrita.edu

M. Anand Kumar
e-mail: m_anandkumar@cb.amrita.edu

K. P. Soman
e-mail: kp_soman@amrita.edu

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_90

ing the closed-domain relation extraction using deep learning methods. In closed-domain relation extraction, the number of entities and the relation between them are fixed. Previous machine learning based information extraction system employs handcrafted features and uses existing Natural Language Processing (NLP) pipelines. These crafted features may fail to capture all the relevant information. The errors present in the pre-existing pipelines will also propagate through the system which hinders the performance of these systems. The deep learning based approach for relation extraction is resilient and more independent of manual feature engineering, thereby relying on pre-existing NLP pipelines decreases.

2 Background

Relation extraction is a key task for building knowledge systems, extending existing ontology like BabelNet, and acting as a knowledge base for question answering systems. Relation extraction is done via various methods like hand-built patterns, bootstrapping with a small highly annotated data, distant supervision method which efficiently uses the existing knowledge bases like freebase, wikidata to extract relations on a bigger corpus [2], and supervised methods. In supervised methods, getting a large quantity of annotated data for the training is key for the high performance of the model. In this work, we are focused toward the supervised relation extraction methods.

2.1 Convolutional Neural Network

The convolutional neural network (CNN) is a kind of feed-forward neural network which is obtained by a slight variation from the multilayer perceptron. In CNN, in order to retain the spatial information of the image, convolution is performed over the image with the different sized filters instead of connecting each unit in the hidden layer to all the units in the previous layer [3]. After convolution operation, a nonlinear activation function like tanh squashes each element to a value between -1 and 1. After applying nonlinearity, a max pooling is done over the output with a fixed size filter which downsamples the information along the spatial dimension. The max pooling operation is done with an intent to capture the most prominent feature in the image, in spite of the slight variation of the spatial features caused by rotation, scaling, and translation. In general, a convolutional neural network will have several stacks of convolution followed by a nonlinear activation before making it to fully connected layer for classification. Several convolutional layers are generally used since the filters in the initial layer learn to detect edge features in the images and the later layers use these low-level features to learn higher level features of images.

2.2 Temporal Convolutional Neural Network

The phenomenal success of the convolutional neural network in the image domain has inspired many researchers to experiment CNN in various domains. In text domain, in order to best understand the meaning, it is necessary to consider the text as a temporal sequence (i.e., each word occurring in the space of time) rather than considering them as an independent feature. The convolution is seen as the operation of blending of features, and it is inferred that convoluted representation has information of nearby words. In NLP, operations are performed over representations given to the text. These representations are given either in the word or character level. In word level, each word is given a vector representation, which is either randomly initialized or assigned through the distributed representations of words [4]. Vector dimension or length of the representations is empirically fixed; this corresponds to the amount of information contained in it. In temporal convolution, the convolution is performed over the temporal sequence [5]. Here, the filter dimension must be equal to the vector dimension of the word embedding. So during convolution, filters slide over the words. In convolution operation, over a window of h words, a filter w is applied to produce new features.

$$o_i = f(w.x_{i:i+h-1} + b) \tag{1}$$

Here, o_i is new feature obtained for the words $x_{i:i+h-1}$, b is the bias term, and f is a nonlinear function like Relu. Similarly, features are obtained for all possible word windows.

$$o = \{o_1, o_2, \ldots, o_{i:i+h-1}\} \tag{2}$$

Then max-over time pooling is done over every feature map of all the available filter, which gives $\hat{o} = \max\{o\}$. $\hat{o}$ is the feature of the corresponding filter. The idea of max-over time pooling is to capture the most significant feature in each feature map. The max-over time pooling naturally deals with different length sequences and decreases the temporal dimension. The pooling operation is generally connected to a dense layer and output layer. The number of classes is equal to the number of neurons in the output layer.

3 Relation Extraction Corpus

In this work, we have used two openly available corpora for relation extraction. The first corpus is taken from SemEval-2010 Task-8 is multi-way classification of semantic relations between pairs of nominals [6], and the second corpus taken from the SemEval-2018 Task-7 is semantic relation extraction and classification in scientific papers [7]. The number of training instances, number of relation classes, and

Table 1 My caption

Sl. No	Corpus	No of instances train–test split	No. of relation classes
1	SemEval-2010 task-8 10717 sentences	75–25	10
2	SemEval-2018 task-7 350 documents	80–20	6

Table 2 SemEval-2010 corpus statistics

Sl. No	Relation	Freq (%)
1	Entity–origin	9.1
2	Instrument–agency	6.2
3	Product–producer	8.8
4	Component–whole	11.7
5	Entity–destination	10.6
6	Cause–effect	12.4
7	Message–topic	8.4
8	Content–container	6.8
9	Member–collection	8.6
10	Other	17.4

Table 3 SemEval-2018 corpus statistics

Sl. No	Relation type	Example	Freq (%)
1	Usage	x is used for y	39
2	Part–whole	x is found in y	19
3	Result	x gives result as y	5.8
4	Topic	x deal with topic y	1.4
5	Compare	x is compared to y	7.7
6	Model	x is model of y	26.5

train–test split of the two corpora are given in Table 1. The class distribution frequency of the two corpora is given in Tables 2 and 3. For the SemEval-2010 corpus shown in Table 2, the frequency of class distribution except "other" class is pretty evenly distributed. For the SemEval-2018 corpus shown in Table 3, the "usage", "model", and "part–whole" have more occurrences when compared with the rest of the other classes.

A preprocessed sample instance from the two corpora is given below: SemEval-2010 task-8 corpus

The (e1) cancer (e1) was caused due to radiation (e2) exposure (e2).

SemEval-2018 task-7 corpus

The key features of the system include (i) robust efficient (e1) parsing (e1) of (e2) korean (e2) (e3) verb final language (e3) with (e4) overt case markers (e4), relatively (e5) free word order (e5), and frequent omissions of (e6) arguments (e6); (ii) high-quality (e7) translations (e7) via (e8) word sense disambiguation (e8) and accurate (e9) word order generation (e9) of the (e10) target language (e10); and (iii) (e11) rapid system development (e11) and porting to new (e12) domains (e12) via (e13) knowledge-based automated acquisition of grammars (e13).

4 Methodology

In this approach, the closed-domain relation extraction is a problem that is modeled as a multi-class classification problem. Some of the constraints of relation extraction are loosened since the entities in the text are already known. In this approach, temporal convolution neural network is employed for relation extraction. The temporal convolution neural network typically models the sequence information present in the text but in order for convolution network to be aware of the entities in sentence or context, the position information about the entities are fed into the network [8]. This is done by feeding an additional array per entity into the network along with the word embedding information as shown in Fig. 1. These arrays have the distance information of each word with respect to the index of the entities.

For a cause–effect relation example as shown in Fig. 1, ***e*1** is the effect and ***e*2** is the cause. The position array of the first entity $e1 = [-1, 0, 1, 2, 3, 4, 5, 6]$ and the second entity $e2 = [-7, -6, -5, -4, -3, -2, -1, 0]$ is formed; this is done by calculating the distance of each word from the entity. The length of position array will be equal to the length of sentence. While training, the sequence is sufficiently zero padded wherever necessary. To avoid the negative values in the position array, it is converted into an index which starts from zero. The position indices along with the word representation are fed into the temporal convolutional neural network. The word

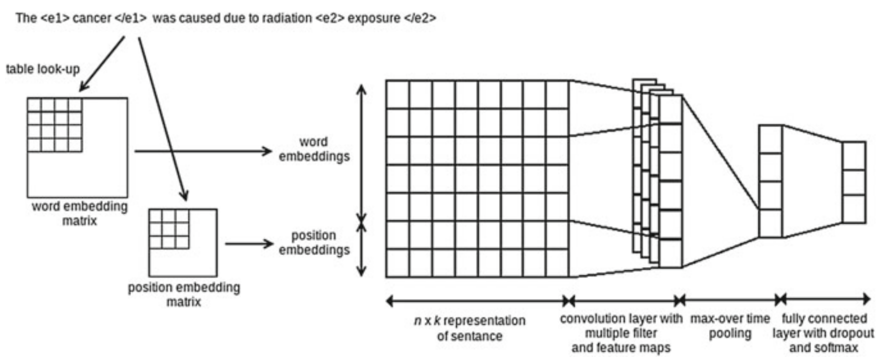


Fig. 1 CNN-based relation extraction architecture

representation is initialized with the word vector from distributed representation [9] word embedding models like word2vec [10] and GloVe [11].

The three variants of temporal CNN used for relation extraction are listed below.

Temporal CNN: This variant uses the word representation initialized through the distributed representation of words by using popular distributed representation framework like word2vec and position embedding information as features.
Temporal CNN with word channel: In this variant in addition to the word and position embedding features, a separate temporal CNN is trained on the word; this forms the local representation for the words with respect to the sentences. This channel is called as word channel and is concatenated with the temporal CNN with word and position embedding features after max-over time pooling. Note that temporal CNN's word embeddings are initialized from the pretrained word embeddings while word channel embeddings are locally learned.
Multichannel temporal CNN: This variant uses two channels of temporal CNN where each channel is initialized with the word vector from different distributed representation frameworks like word2vec and GloVe. Then these two channels are concatenated before max-over time pooling. This idea this model is inspired by the work [12].

5 Result and Discussion

In this work, temporal convolutional neural network based relation extraction model is applied on SemEval-2018 task-7 and SemEval-2010 task-8 corpus. On these two corpora, we have applied three variants of temporal CNN. The first model is a temporal CNN architecture with pretrained word embeddings and position embedding as features; this is considered as the baseline. In the second model, in addition to a temporal CNN with word and position embedding, an additional word channel is introduced; it is seen that having an additional word channel improves the accuracy and F1 score of the classification task. Then, in the final model, temporal CNN with multichannel word embeddings is used. It is observed that using different distributed word representation frameworks significantly helps the model to learn better representation for sentences, thereby increasing the overall performance of the classification task. Tables 4 and 5 show the accuracy and F1 score of our various approaches on the two corpora. So, it is seen that model which uses additional word representation channel improves the overall accuracy of the classification task. The state-of-the-art (SOTA) performance reported for the SemEval-2010 task-8 corpus is using bidirectional gated recurrent unit with attention [13].

Table 4 Results for SemEval-2010 corpus

Sl. No	Architecture	Filter length	Accuracy	F1 (macro)
1	Temporal CNN	3	0.77	0.740
2	Temporal CNN	4	0.76	0.741
3	Temporal CNN with word channel	3	0.78	0.761
4	Temporal CNN with word channel	4	0.78	0.764
5	Multichannel temporal CNN	3	0.79	0.769
6	Multichannel temporal CNN	4	0.79	0.771
7	BiGRU with ATT (SOTA)	–	–	0.84

Table 5 Results for SemEval-2018 corpus

Sl. No	Architecture	Filter size	Accuracy	F1 (macro)
1	Temporal CNN	3	0.5512	0.291
2	Temporal CNN	4	0.5441	0.280
3	Temporal CNN with word channel	3	0.5657	0.327
4	Temporal CNN with word channel	4	0.5749	0.316
5	Multichannel temporal CNN	3	0.5896	0.325
6	Multichannel temporal CNN	4	0.6196	0.313

6 Conclusion and Future Work

In this paper, the task of relation extraction is modeled as a multi-class classification problem. For classifying relations, temporal convolution neural network is employed and the word representation is done using distributed representation framework like word2vec and GloVe. In this work, three variants based on temporal convolutional neural network are applied on SemEval 2018 and 2010 relation extraction corpus and it is observed that model which has multiple channels of word representation helps and improves the overall accuracy of the relation extraction task. For the future work, we are planning to implement a multi-level attention model which uses attention to learn a better representation for the words and also employs attention mechanism for classification. In our next work, we are planning to make an end-to-end network which does named entity recognition and relation extraction together.

References

1. Devi GR, Veena PV, Anand Kumar M, Soman KP (2018) Entity extraction of Hindi–English and Tamil–English code-mixed social media text. In: Forum for information retrieval evaluation. Springer, pp 206–218
2. Mintz M, Bills S, Snow R, Jurafsky D (2009) Distant supervision for relation extraction without labeled data. In: Proceedings of the joint conference of the 47th annual meeting of the ACL and the 4th international joint conference on natural language processing of the AFNLP, vol 2, pp 1003–1011. Association for Computational Linguistics
3. LeCun Yann, Bengio Yoshua et al (1995) Convolutional networks for images, speech, and time series. Handb Brain Theor Neural Netw 3361(10):1995
4. Barathi Ganesh HB, Anand Kumar M, Soman KP (2018) From vector space models to vector space models of semantics. In: Forum for information retrieval evaluation. Springer, pp 50–60
5. Kim Y (2014) Convolutional neural networks for sentence classification. arXiv preprint arXiv:1408.5882
6. Hendrickx I, Kim SN, Kozareva Z, Nakov P, Séaghdha DÓ, Padó S, Pennacchiotti M, Romano L, Szpakowicz S (2009) Semeval-2010 task 8: multi-way classification of semantic relations between pairs of nominals. In: Proceedings of the workshop on semantic evaluations: recent achievements and future directions, pp 94–99. Association for Computational Linguistics
7. Ga´bor K, Buscaldi D, Schumann A-K, QasemiZadeh B, Zargayouna H, Charnois T (2018) Semeval-2018 Task 7: semantic relation extraction and classification in scientific papers. In: Proceedings of international workshop on semantic evaluation (SemEval-2018), New Orleans, LA, USA
8. Zeng D, Liu K, Lai S, Zhou G, Zhao J (2014) Relation classification via convolutional deep neural network. In: Proceedings of COLING 2014, the 25th international conference on computational linguistics: technical papers, pp 2335–2344
9. Barathi Ganesh HB, Anand Kumar M, Soman KP (2016) Distributional semantic representation for information retrieval. In: Working notes of FIRE, pp 1–7
10. Mikolov T, Sutskever I, Chen K, Corrado GS, Dean J (2013) Distributed representations of words and phrases and their compositionality. In: Advances in neural information processing systems, pp 3111–3119
11. Pennington J, Socher R, Manning C (2014) Glove: global vectors for word representation. In: Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP), pp 1532–1543
12. Shi H, Ushio T, Endo M, Yamagami K, Horii N (2016) A multichannel convolutional neural network for cross-language dialog state tracking. In: 2016 IEEE Spoken language technology workshop (SLT). IEEE, pp 559–564
13. Kim J, Lee J-H (2017) Multiple range-restricted bidirectional gated recurrent units with attention for relation classification. arXiv preprint arXiv:1707.01265

Multi-Object Detection Using Modified GMM-Based Background Subtraction Technique

Rohini Chavan, S. R. Gengaje and Shilpa Gaikwad

Abstract Detection of objects is the most important and challenging task in video surveillance system in order to track the object and to determine meaningful and suspicious activities in outdoor environment. In this paper, we have implemented novel approach as modified Gaussian mixture model (GMM) based object detection technique. The object detection performance is improved compared to original GMM by adaptively tuning its parameters to deal with the dynamic changes that occurred in the scene in outdoor environment. Proposed adaptive tuning approach significantly reduces the overload experimentations and minimizes the errors that occurred in empirical tuning traditional GMM technique. The performance of the proposed system is evaluated using open source database consisting of seven video sequences of critical background condition.

1 Introduction

Human requirement of automated detection system in personal, commercial, industrial, and military areas leads to development of video analytics which will make lives easier and enable us to compete with future technologies [1]. On the other hand, it pushes us to analyze the challenges of automated video surveillance scenarios. Humans have an amazing capacity for decision-making but are notoriously poor at maintaining concentration levels. A variety of studies has shown that after 20 min of watching, up to 90% of the information being shown on monitors will be missed. As closed-circuit television (CCTV) culture continues to grow, humans

R. Chavan (✉) · S. Gaikwad
Bharati Vidyapeeth University College of Engineering, Pune, India
e-mail: chavanrohini10@gmail.com

S. Gaikwad
e-mail: spgaikwad@bvucoep.edu.in

S. R. Gengaje
Walchand Institute of Technology, Solapur, India
e-mail: srgengaje@rediffmail.com

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_91

would require to observe feed from hundreds of camera 24×7 [2]. It shows requirement of automatic system that analyzes and stores video from 100s of cameras and other sensors, detecting events of interest continuously and browsing of data through sophisticated user interface. It is simply known as video analytics [3, 4].

Recent research in computer vision is giving more stress on developing system for monitoring and detecting humans. It is helpful for people in personal, industrial, commercial, and military areas to develop innovations in video analysis, to compete with future technologies and to accept the challenges in automatic video surveillance system. Video surveillance tries to detect, classify, and track objects over a sequence of images and help to understand and describe the object behavior by human operator. This system monitors sensitive areas such as airport, bank, parking lots, and country borders. The processing framework of an automated video surveillance system includes stages like object detection, object classification, and object tracking. Almost every video surveillance system starts with motion detection. Motion detection aims at segmenting regions of interest corresponding to moving objects from remaining image. Subsequent processes such as object classification and tracking performances are greatly dependent on it. If there is significant fluctuations in color, shape, and texture of moving object, it causes difficulty in handling these objects. A frame of video sequence consists of two groups of pixels. The first group represents foreground objects and second group belongs to background pixels. Different techniques such as frame differencing, adaptive median filtering, and background subtraction are used for extraction of objects from stationary background [5]. The most popular and commonly used approach for detection of foreground objects is background subtraction. The important steps in background subtraction algorithm are background modeling and foreground detection [6]. Background modeling gives reference frame which represents statistical description of entire background scene. The background is modeled to extract interested object from video frames. It is designed with first few frames of video sequence. But, in case of quasi-stationary background such as wavering of trees, flags, and water, it is more challenging to extract exact moving object. In this situation, single model background frame is not enough to accurately detect the moving object but adaptive background modeling technique is used for exact detection of objects from dynamic background [7].

2 Object Detection Using Adaptive Gaussian Mixture Model

2.1 Basic Gaussian Mixture Model

A Gaussian mixture model (GMM) is parametric probability density function presented as a weighted sum of K Gaussian component densities [8, 9]. It is represented by the following equation:

$$P(x_t) = \sum_{i=1}^{k} \left(\omega_{\mathrm{i,t}}\, n(X_t; \mu_{i,t}, \sum i, t) \right) \tag{1}$$

$$\sum_{i=1}^{k} \left(\omega_{\mathrm{i,t}} \right) = 1 \tag{2}$$

where x is a D dimensional data vector and ω is weight of i^{th} Gaussian component. Here, k is the number of Gaussian distributions, t represents time, μ is mean value of the ith Gaussian mixture at time t, and $\sum i, t$ is the covariance matrix. The entire GMM is scaled by mean vectors, covariance matrices, and mixture weights of all component densities. The mean of such mixture is represented by following equation:

$$\mu_t = \sum_{i=1}^{k} \omega_{i,t} \mu_{i,t} \tag{3}$$

There are several variants on the GMM and covariance matrices constrained to be diagonal. The selection of number of components and full or diagonal covariance matrix is often determined by the availability of data for estimating GMM parameters.

Background subtraction object detection technique is popular as it is less complex, simple and easy to implement. It takes the difference between current frame (I_t) and reference frame. The reference frame is denoted by (B_{t-1}). Hence, difference image (D_t) is given by

$$D_t = |B_{t-1} - I_t| \tag{4}$$

Foreground mask (F_t) is given by applying threshold to difference image

$$F_t = 1, \text{ when } D_t > \mathrm{Th}$$
$$F_t = 0, \text{ when } D_t < \mathrm{Th}$$

2.2 GMM Model Initialization and Maintenance

For stationary process pixels, EM algorithm is applicable. K-means algorithm is an alternative to EM [10]. Using K-means approximation, every new pixel value X_t is checked against existing K Gaussian distribution until match is found. A match is given by

$$\mathrm{Sqrt}\left(\left(X_{t+1} - \mu_{i,t} \right) T. \sum\nolimits_{i,t}^{-1} \left(X_{t+1} - \mu_{i,t} \right) \right) < k\sigma_{i,t} \tag{5}$$

where k is constant threshold value which is selected as 2.5. If K distribution is not matched with current pixel value then least probable distribution is replaced with current distribution value as its mean, weight, and variance. Prior weights of K distributions at time t are adjusted as follows:

$$\omega_{k,t} = (1 - \alpha)\omega_{k,t} - 1 + \alpha(M_{k,t}) \tag{6}$$

where α is learning rate and $M_{k,t}$ is 1 for model which is matched and 0 for other models. After operating this approximation, weights are again normalized. The μ and σ parameters remain same for unmatched distributions. The parameters of distribution which matches new observations are updated as follows:

$$\mu_{\mathrm{t}} = (1 - \rho)\mu_{\mathrm{t}} - 1 + \rho X_t \tag{7}$$

$$\sigma t = (1 - \rho)\sigma 2t - 1 + \rho(X_{\mathrm{t}} - \mu t)T(X_t - \mu t) \tag{8}$$

$$\sigma_t = (1 - \rho)\sigma_{\mathrm{t}-1} + \rho(X_{\mathrm{t}} - \mu t)^{\mathrm{T}}(X_{\mathrm{t}} - \mu t) \tag{9}$$

where

$$\rho = \alpha.n(X_{\mathrm{t}}|\mu_{\mathrm{k}}, \sigma) \tag{10}$$

One advantage of this technique is that when new thing is added in the model then it will not completely destroy the previous background model but it can update the model.

3 Object Detection Using Adaptive GMM

The system is implemented using two steps such as GMM-based object detection and noise removal using morphological operations. Implementation is done using MATLAB 2014v with the help of computer vision system toolbox. The small detected regions whose area is less than moving object and which are not part of foreground object can be removed using noise removal algorithm. Finally, output binary image is compared with ground truth image for performance evaluation to determine accuracy.

Background modeling is adaptive to accommodate all the changes occurring in the background scene. It is very sensitive to dynamic changes that have occurred in the scene which causes consequent need of adaptation of background as per the variations in background. The research has progressed toward improving robustness and accuracy in background subtraction method for complex background condition like sudden and slow illumination change. A common attribute of BS algorithm is learning rate, threshold, and constant parameter K which can be empirically adjusted to get desired accuracy. However, tuning process for these parameters has been less attentive due to lack of awareness. Stauffer and Grimson [7] suggested that selection of learning rate and threshold value is important among all other parameters. Tuning process for these parameters requires time intense repeated experimentation to achieve optimum results. It is very challenging to set the parameters because it requires understanding of background situation and common setting for different scenarios may not produce accurate result. All these aspects put limitations on effective use of background subtraction algorithm and demand improvement and extension of original GMM.

Recent years, researchers are focused on developing innovative technology to improve performance of IVS in terms of accuracy, speed, and complexity. To design novel approach for GMM parameter tuning based on extraction of statistical features and map with GMM training parameters [11]. Learning rate parameter is very important which determines the rate of change of background. Large amount of experimentation is required to set the value of learning rate for exact detection of foreground object. It is required to develop the system which tunes the parameter automatically for satisfactory performance of GMM.

GMM modeling is able to handle multimodal background scene. Performance of GMM-based background subtraction is decided by pixel-wise comparison of ground truth and actual foreground mask. Performance of the system is evaluated with the help of primary metrics such as true positive (TP), true negative (TN), false positive (FP), and false negative (FN) and secondary metrics like sensitivity, accuracy, miss rate, recall, and precision. Precision reflects false detection rate and recall gives accuracy of detection. Precision and recall are the two important measures in order to estimate detection algorithm systematically and quantitatively [12, 13].

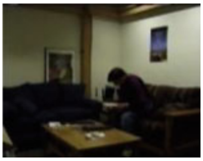
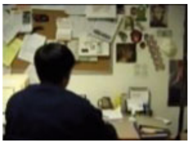
$$\text{Precision}(\%) = \frac{\text{TP}}{\text{TP} + \text{FP}} * 100 \tag{11}$$

Our proposed system brings innovation in original GMM-based object detection system through tuning and adaptation of important parameters such as number of component, learning rate, and threshold.

3.1 Video Database

Wallflower database is open source database [9]. It includes seven sets of video sequence with different critical situations in background. Video frames of size 160 × 120 pixel, sampled at 4 Hz. Data set provider also gives one ground truth image and text file having description of all video sequences. Ground truth is binary image representing foreground mask of specific frame in video sequence. Table shows all test sequences along with their ground truth (Table 1).

Table 1 Wallflower dataset of seven different video sequences along with ground truth

Sr. No.	Name	Test sequence	Ground truth
1	Moved object (MO)		
2	Time of day (TOD)		
3	Light switch (LS)		
4	Waving tree (WT)		
5	Camouflage (C)		
6	Foreground aperture (FA)		
7	Bootstrap (B)		

3.2 Experimental Setup

The main focus of research is based on appropriate selection of GMM training parameters like K, α, and T. The selection of K Gaussian component value is function of complexity of background scene. If the background is simple and unimodal, the value of K must be selected as 1 or 2. For complex multimodal background, value of K is more than 2 and less than 5 so as to improve the accuracy in detection process. Various pairs of α and T are evaluated on Wallflower dataset. After lots of experimentation best pair of α and T is identified based on performance analysis on various Wallflower videos [9]. Parameter initialization, training, and testing are the three important steps for object detection process.

3.3 Parameter Initialization

Object detection system includes various GMM parameters like number of training frames, initial variance, and training parameters (K, α, and T). They are initialized as follows:

Number of training frames: 200 (given by data set provider),
Number of component: 4,
Initial variance: 0.006, and
Threshold: Adjusting value empirically (0.5, 0.6, 0.7, 0.8, 0.9).

4 Experimental Results

GMM-based object detection system is evaluated using various settings of α and T for each sequence. After this experimentation, for all videos, appropriate setting of α and T is decided based on lowest value of total error. Performance metrics are calculated for each sequence by comparing detected mask with ground truth.

Results are as follows (Fig. 1):

GMM-based background subtraction technique gives best overall detection performance at $\alpha = 0.001$ and $T = 0.9$. These parameter settings improve the accuracy of foreground mask which is almost matching with ground truth. Learning rate and threshold have enough power to tune object detection performance. Best overall performance setting has less probability to give best result at individual level. Best individual performance may be obtained by different settings of parameters for some of the sequences. Performance analysis for best α and T can be done at pixel level. Empirically selection of higher threshold value gives merging of foreground objects with background. It leads to increase in false negative and decrease in true positive. For faster changing background, empirical selection of lower value of α is too low to adapt such background changes. Empirical setting of threshold value, $T = 0.9$, is

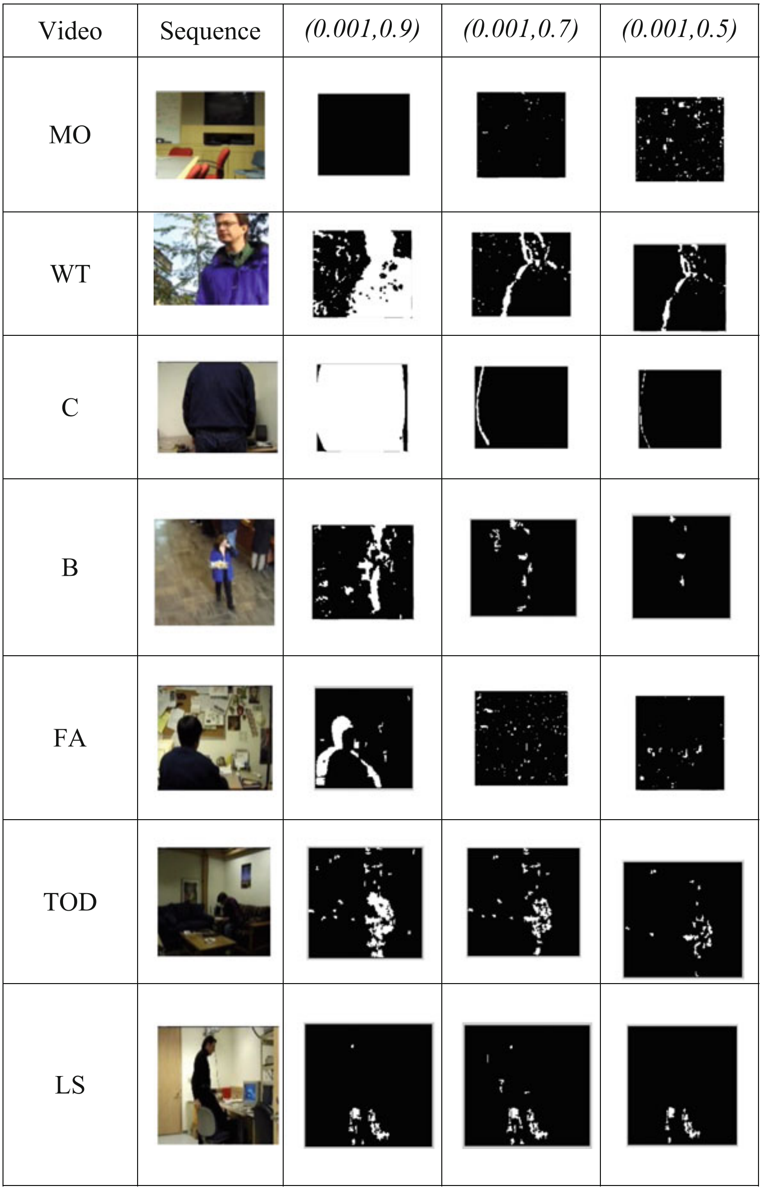


Fig. 1 Foreground mask obtained using GMM for different values of α and T

being so high that all foreground pixels are merged into the background. It gives an increase in false negative and decrease in true positive pixels. Same way, empirical selection of learning rate, $\alpha = 0.001$, is being too low for rapid changing background. Thus, misclassification is higher and accuracy is lower for those video sequences in

Table 2 Performance evaluation of proposed system on Wallflower dataset

Video	True positive (TP)	True negative (TN)	False positive (FP)	False negative (FN)	Precision (%)	Accuracy (%)
MO	0	19,200	0	0	100	100
WT	12,558	766	306	4570	90	80.34
C	3399	1039	770	8231	81.53	83.68
B	2546	15,927	295	432	89.61	85.49
TOD	125	17,762	0	1313	90.16	91.43
FA	3754	13,718	745	983	83.44	84.69
LS	718	12,670	3359	2453	69.72	77.87

which sudden change of illumination is occurring. This experimentation also suggests that various settings of (α and T) may result in more improved performance for different video sequences than fixed selection of (α and T) (Table 2).

5 Conclusion

Proposed research emphasizes on proper tuning of important GMM parameter leading to improvement in the performance accuracy of GMM-based object detection system. GMM parameters mainly include number of mixture component (K), learning rate (α), and threshold (T). We have implemented two approaches for tuning these parameters such as traditional empirical tuning and automated adaptive tuning based on background dynamics. Traditional empirical tuning method is implemented using different settings of α and T, while K is kept constant to high value for complex scene. After large number of experimentation, appropriate pair of α and T is selected based on low performance error.

Proposed adaptive tuning method involves adaptation of α and keeping T and K constant to appropriate value. Unique EIR concept is used to extract background dynamics for current frame. Learning rate is tuned depending on EIR. This modified approach improves the result of GMM compared to the original GMM. This result strongly emphasizes the strength of learning rate adaptation. Performances of GMM with these tuning methods are evaluated based on foreground mask obtained using GMM and ground truth image in database. The analysis of performance can be done using primary metrics such as TP, TN, FP, and FN as well as secondary metrics like precision and accuracy. Our proposed system performance is compared with traditional empirical method and other existing techniques. Our research is implemented on MATLAB 2014 platform. Different functions from MATLAB computer vision toolbox are used for implementation of algorithm.

References

1. Hsieh J-W, Yu S-H, Chen Y-S (2006) An automatic traffic Surveillance system for vehicle tracking and classification. IEEE Trans Intell Transp Syst 7
2. Chauhan AK, Krishan P (2013) Moving object tracking using Gaussian mixture model and optical flow. Int J Adv Res Comput Sci Softw Eng 3
3. Picardi M (2004) Background subtraction technique: a review. In: IEEE international conference on systems, man and cybermetics
4. Viola P, Jones M (2005) Detecting Pedestrians using patterns of motion and appearance. Int J Comput Vision 63(2):153–161
5. Bo W, Nevatia R (2007) Detection and tracking of multiple partially occluded Humans by Baysian combination of edgelet based part detector. Int J Comput Vision 75(2):247–266
6. Ran Y, Weiss I (2007) Pedestrian Detecting via periodic motion analysis. Int J Comput Vision 71(2):143–160
7. Stauffer C, Grimson WEL (1999) Adaptive background mixture models for real time tracking. In: International conference on computer vision and pattern recognition 2
8. Cucchiaria R, Grana C, Piccardi M, Prati A (2003) Detecting moving objects, ghosts and shadows in video streams. IEEE Trans PAMI 25(10):1337–1342
9. Toyama K, Krumm J, Brumitt B (1999) Wallflower: Principles and practice of background maintenance. In: International conference of computer vision, pp 255–261
10. Zhang, LZ, Hou Z, Wang H, Tan M (2005) An adaptive mixture Gaussian background model with online background reconstruction and motion segmentation. ICIT, pp 23–27
11. White B, Shah M (2007) Automatically tuning background subtraction parameters using Particle swarm optimization. In: IEEE international conference on multimedia and Expo, China, pp 1826–1829
12. Harville M, Gordon G, Woodfill J (2001) Foreground segmentation using adaptive mixture models in color and depth. In: Proceeding of the IEEE workshop on detection and recognition of events in Video, Canada
13. Elgammal A, Harwood D, Davis L (2000) Non parametric model for background subtraction. In: European conference on computer vision, pp 751–767

Motion Detection Algorithm for Surveillance Videos

M. Srenithi and P. N. Kumar

Abstract Locality Sensitive Hashing (LSH) is an approach which is extensively used for comparing document similarity. In our work, this technique is incorporated in a video environment for finding dissimilarity between the frames in the video so as to detect motion. This has been implemented for a single point camera archiving, wherein the images are converted into pixel file using a rasterization procedure. Pixels are then tokenized and hashed using minhashing procedure which employs a randomized algorithm to quickly estimate the Jaccard similarity. LSH finds the dissimilarity among the frames in the video by breaking the minhashes into a series of band comprising of rows. The proposed procedure is implemented on multiple datasets, and from the experimental analysis, we infer that it is capable of isolating the motions in a video file.

1 Introduction

Smart video surveillance plays a vital role in ensuring safety and security in public and private spaces. Particularly in vast spaces and big organizations, possession of an inherent capability to detect motion without human intervention assumes importance (where large number of cameras may be deployed). The pertinent challenging factors in the detection of motion from surveillance videos are fatigue for a human to review the video; manifold increase in computation cost for further processing and the requirement of large storage space. Thus, as a part of any intelligent video surveillance system, inclusion of video motion detection capability that can automatically detect the events and render further support in the ensuing video processing techniques, is a force multiplier.

M. Srenithi · P. N. Kumar (✉)
Department of Computer Science and Engineering, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Coimbatore, India
e-mail: pn_kumar@cb.amrita.edu

M. Srenithi
e-mail: cb.en.p2cse16018@cb.students.amrita.edu

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_92

2 Related Work

Traditional motion detection approaches are background subtraction; frame differencing; temporal differencing and optical flow. Background subtraction method detects motion by pixel-by-pixel subtraction of the current image from a reference background image [1]. Frame differencing identifies the existence of moving object by considering the variation between two successive frames [2]. In these methods, the performance is adversely affected when the background is dynamic. In temporal differencing [3], the current frame is compared with the previous frame, and the compared frame is then threshold to fragment out the foreground objects [4]. This approach fails when the pixel values are uniformly distributed. Optical flow [5] is the motion of objects in a visual sight caused by relative motion between the camera and the frame. This method, however, is computationally expensive. Gregory presented a content-based retrieval method for long-surveillance videos in which archival of video is based on low-level spatio-temporal extraction [5]. They are hashed into an inverted index using Locality Sensitive Hashing (LSH) for query flexibility. Partial matches are extracted to user-created queries and [5] assembled subsequently into full matches using dynamic programming that assembles the indexed low-level features into a video segment that matches the query route by exploiting causality [5]. Ling et al. [6] presented a contour based object tracking approach. Using multi-feature fusion strategy, the rough location of an object is found. Extracting the contours with the help of region-based object contour extraction, accurate and robust object contour tracking have been demonstrated. Zhao et al. [7] presented a method in which the average values of a continuous multi-frame gray image are calculated, and through statistical averaging of the continuous image sequence the background image is obtained.

3 Background

Hashing has gained much popularity for quick estimate of similarity computation. Among the diverse hashing methods, Locality Sensitive Hashing is extensively used in high dimensional databases. In LSH [8], a group of arbitrary hash functions called Minhash are associated with the banding technique to eliminate pairs with high similarity from the pairs with less similarity. Minhash transforms large files into a matrix of signatures which preserves the similarity of record pairs and LSH finds identical pairs which are most probably similar in the same bucket, thus abstaining assimilation of each and every combination of files.

3.1 Minhashing

Minhashing is a technique which uses the random hash function for fast Jaccard similarity computation. Amid Minhash and Jaccard Similarity there exists a significant relationship, which is the probability that the Jaccard similarity for two sets equals the value given by the Minhash function for those sets. To apprehend how, it is essential to visualize the columns of those sets *A* and *B*, therefore rows are separated into three categories: category *X* rows contain 1 in both columns, category *Y* rows contain 1 and 0 alternatively, and category *Z* rows contain 0 in both. Meanwhile, most of the rows are of category *Z*; Jaccard similarity (*A*, *B*) is determined by the ratio of category *X* and *Y* [9]. Thus Jaccard similarity $(A, B) = x/x+y$ where *x* denotes rows of category *X* and *y* denotes rows of category *Y*. As the size of $A \cap B$ is *x* and the size of $A \cup B$ is $x + y$ [9].

Assume for probability $h(A) = h(B)$ that the rows are randomly permuted, and when we traverse from the top $x/(x + y)$ is probability to see category *X* before a category *Y*. When the first row from the top is a category *X*, then $h(A) = h(B)$ [9]. Otherwise, if the first row is a category *Y*, then the Minhash value is acquired by set with a 1. Thus, $h(A) \neq h(B)$ if category *Y* row is meet first [9]. Thus concluding, the probability $h(A) = h(B)$ is $x/(x + y)$, which is the Jaccard similarity of *A* and *B*.

3.2 Minhash Signatures

To represent sets as a characteristic matrix *M*, a random n number of permutations of the rows of the matrix are chosen. Minhash functions are determined as $h_1, h_2, \ldots, h_n$. The minhash signature for *S* is constructed from the column representing set S, that is the $[h_1(S), h_2(S), \ldots, h_n(S)]$ vector [9]. This hash-values list is represented as a column [9]. Thus, a signature matrix is formed from matrix *M*, in which the minhash signature replaces the ith column of M for the set of the ith column.

3.3 Locality Sensitive Hashing

Although the concept of minhash function can be employed to compress bulk files into small signatures and conserve the required similarity, efficient similarity computation is still impossible [9]. Since the total number of pairs of files will be too large. Hence, minhash is associated with the banding technique Locality Sensitive Hashing to facilitate efficient similarity computation. In Locality Sensitive Hashing, files are hashed numerous times, and such that similar files are probably hashed to the same bucket than dissimilar files are. Any pair of files that is hashed to the same bucket for any hash function becomes a candidate pair. Candidate pairs are considered for similarity. This intends that dissimilar pairs by no means hash to the same bucket.

Having the minhash signatures, an efficient way to further handle the hashing is to divide the signatures into series of bands b comprising of rows r [9]. There is a hash function for each band that takes a column's segment inside that band and hashes them to some extensive number of buckets. The equivalent hash function can be utilized for all the bands, but for each band we use a separate bucket array, so columns with the same vector in different bands will not hash to the same bucket [9].

4 Proposed Method

In this work, we consider the problem of detecting motion in sequence of frames within a video and propose a procedure based on minhashing and LSH. We use rasterizing procedure [10] in which frames of the video are chosen, raster layer objects created, and finally these raster layer objects are aggregated to store the pixels in a text file. Each frame in the video is modelled to pixel files of type text. In order to detect motion, each pair of frame needs to be compared. This is not only time consuming, but also is prohibitively expensive as well. To surmount this problem we employ minhashing and LSH algorithms [11].

Minhash function tokenizes the pixels into a set of hash integers, from which the minimum value is selected, which is similar to random selection of a token. In LSH, [11] minhashes are shattered into a sequence of bands consisting of series of rows. Each band is hashed to a bucket. If two continuous frames have different minhashes in a band, they are hashed to the same bucket and considered as candidate pairs from which motion is detected. The outline workflow of the proposed method is shown in Fig. 1.

The procedure to implement the LSH algorithm is depicted in the architecture diagram given in Fig. 2.

5 Implementation and Results

5.1 Rasterization of Frames in the Video

Video is considered as frames per second for rasterization and a raster layer object is created from the image object. Raster layer object [10] is a dot matrix data structure, representing a rectangular grid of pixels of a frame as a rectangular grid. The procedure for calculating the pixel values of the frames in the video is as follows:

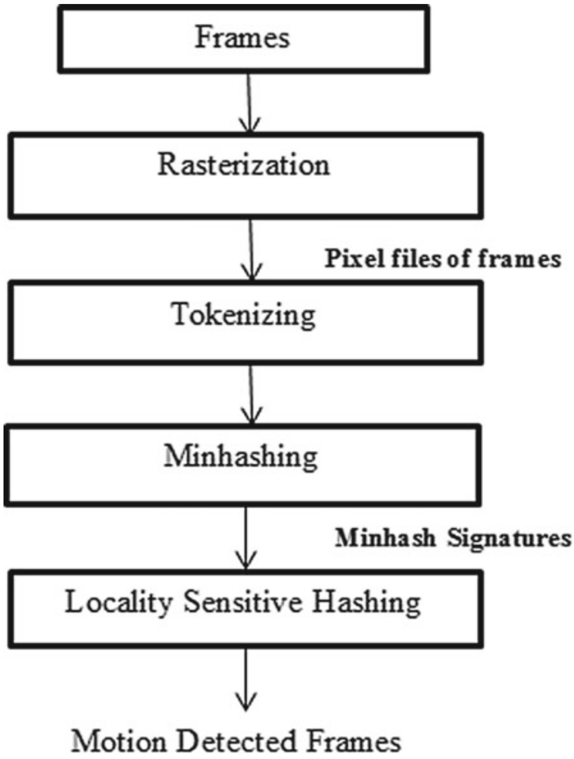


Fig. 1 Proposed workflow of motion detection

Input: Set of frames extracted from the video
Output: Set of text files containing pixel values of each frame.
Steps:

1. Convert each frame as Raster Layer Object.
2. Assign a variable to the extent co-ordinates of frame.
3. Calculate the pixel value of the frames.
4. Pixel values are aggregated and stored in a text file.

5.2 Minhashing

The pixels of the frame are tokenized using a tokenizer, converted into a set of hash integers using a minhash function, and we select the minimum value from the set of integers. This is similar to selecting a random token. Minhash function [11] does this repeatedly by effectively selecting n random token using various hash functions. For random hash functions, minhash generator takes seed for re-creating equivalent

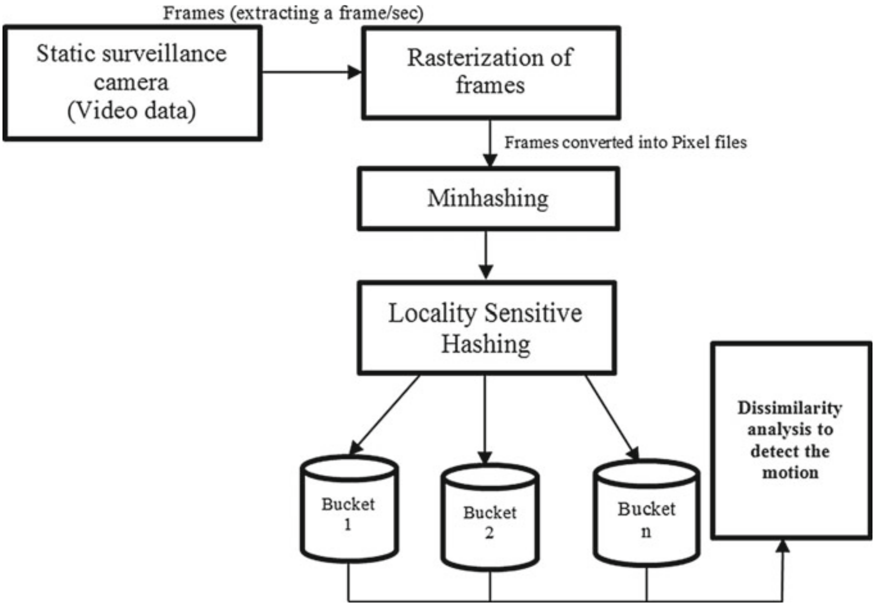


Fig. 2 The LSH implementation architecture

minhash function. Thus minhash function converts the set of tokenized pixels into n randomly selected and hashed tokens.

5.3 Locality Sensitive Hashing

The LSH [11] function solves the problem of complexity. The minhashes are broken down into series of bands comprising of rows. Each band is hashed to a bucket. Frames having exact minhashes in a band are hashed into the different bucket, and frames were motions have occurred that is with different minhashes are hashed into same buckets so as to detect the motions. The procedure for LSH is as follows:

Input: Set of text files containing pixel values of each frame.
Output: Frames containing abnormal events
Steps:

1. Initialize the minhash generator#n and the seed value.
2. Load the pixel files.
3. Tokenize the pixel files.
4. Apply hash functions to convert the tokens into set of hash integers.
5. Select the minimum value token.
6. Break the minhashes into bands and rows.

7. Each band is hashed into buckets.
8. Similar bands are hashed into separate buckets.
9. Dissimilar bands are found in same buckets.

5.4 Datasets

We have considered two datasets for our experiment, viz. (1) A real-time static surveillance dataset and (2) Video Surveillance Online Repository (VISOR). The real-time static surveillance dataset includes approximately one hour long videos captured from a university campus; each consisting of 500 to 600 frames per second with a resolution of 352×288. The video has been captured both during the day (Fig. 3) and at night (Fig. 4). Movement of pedestrians and vehicles has been captured to illustrate motion detection. The VISOR [12] video (Fig. 5) contains one minute long, 60 frames per second data with a resolution of 360×288. It is a static camera traffic surveillance video taken from Video Surveillance Online Repository.

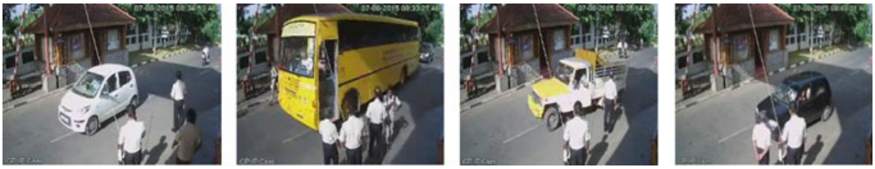


Fig. 3 An illustrative frame of a real-time dataset (Daylight)

Fig. 4 An illustrative frame of a real-time dataset (At night)

Fig. 5 Examples of detected motions in VISOR Dataset

5.5 Results

The results obtained in terms of Precision, Recall and F-Measure employing the min-hash and LSH techniques are tabulated below as a confusion matrix for both data sets used by us. In F-measure twice as much weight is given to precision since accurately detecting the motions is essential, thus an $F_{0.5}$ score is used. In motion detection using this technique, we have achieved an accuracy of over 85% for detection of motion in videos (Table 1).

A comparison of the results achieved in our experiment of employing LSH, with other techniques reported viz., MODE [13], DP-GMM [14], and Culibrk [15] is given in Table 2. From the table, we can infer that the accuracy of the method proposed which is experimented on real-time dataset is significant and is higher than few of the existing methods.

As shown in the Table 3 storing the detected frames reduces the amount of storage up to 90% for real-time and VISOR videos. Also saving the detected frames lessen the fatigue of security personnel to sit and review the surveillance videos.

The Receiver Operating Characteristic (ROC) curve obtained for the motion detection experiment for the real-time static surveillance video and VISOR static surveillance video is as shown in Fig. 6.

Table 1 Results of motion detection employing LSH

Dataset	Precision	Recall	$F_{0.5}$ Measure
Real-time	0.865	0.93	0.88
VISOR	0.84	0.82	0.835

Table 2 Comparison of accuracy obtained/reported

Method	Accuracy reported
Motion detection using LSH	0.87
DP-GMM [14]	0.7567
Culibrk [15]	0.5256
MODE [13]	0.7628
Belhani [16]	0.85

Table 3 Compression ratio of the videos

Video	Actual size (Mb)	Compressed size (Mb)
Video 1 (Real-time)	24.2	2.57
Video 2 (Real-time)	33.3	2.1
Video 3 (Real-time)	27	2.3
VISOR video	15.3	1.5

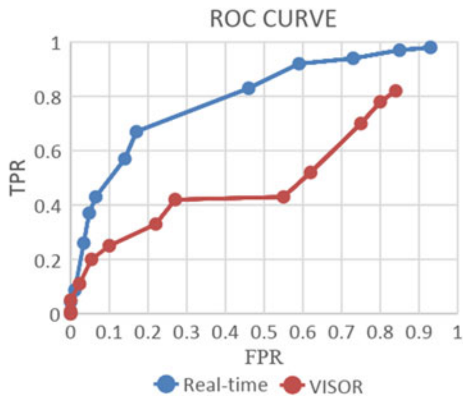


Fig. 6 Receiver operating characteristic of LSH Method

6 Conclusion

In this work, we have developed a motion detection system for static surveillance videos based on minhashing, followed by the application of Locality Sensitive Hashing (LSH) technique. The extracted frames are rasterized and minhashed. The minhashed frames are then broken into a sequence of bands and hashed into buckets employing the LSH technique. Experimental results on the real-time static surveillance and in the VISOR dataset demonstrates the improved accuracy and effectiveness of the proposed method and also out-performs the few of the existing methods. Such a system can be deployed in the monitoring station of surveillance videos in large establishments. This method is not only accurate and efficient but also would reduce the fatigue to the security personnel nominated to screen and monitor large sized recordings of videos in installations and open spaces.

References

1. Joshan J, Suresh P (2012) Systematic survey on object tracking methods in video. Int J Adv Res Comput Eng Technol (IJARCET)
2. Papageorgiou, Oren, Poggio (1998) A general framework for object detection. In: IEEE international conference on computer vision (ICCV)
3. Astha, Manoj, Kailash (2016) Survey on video object detection and tracking. Int J Current Trends Eng Technol
4. Singh T, Sanju, Vijay (2014) A new algorithm designing for detection of moving objects in video. Int J Comput Appl
5. Gregory C, Elgharib M, Saligrama V, Jodoin P-M (2015) Retrieval in long surveillance videos using user-described motion and object attributes. IEEE Trans Circuits Syst Video Technol
6. Lu X, Song L, Yu S, Ling N (2012) Object contour tracking using multi-feature fusion based particle filter. In: IEEE conference on industrial electronics and applications (ICIEA)
7. Yang K, Cai Z, Zhao L (2013) Algorithm research on moving object detection of surveillance video sequence. Opt Photonics J

8. Sowmya K, Kumar PN (2018) Traffic density analysis employing locality sensitive hashing on GPS data and image processing techniques. In: Computational vision and bio inspired computing
9. Rajaraman A, Ullman J (2011) Mining of massive datasets, Cambridge University Press, New York. chap. 3
10. https://www.rdocumentation.org/packages/raster/versions/2.5–8/topics/raster
11. Mullen L Minhash and locality-sensitive hashing. https://cran.r-project.org/web/packages/textreuse/vignettes/textreuse-minhash.html
12. Vezzani R, Cucchiara R (2010) Video surveillance online repository (ViSOR): an integrated framework in Multimedia Tools and Applications, Kluwer Academic Press
13. Singh B, Singh D, Singh G, Sharma N, Sibbal V (2014) Motion detection for video surveillance. In: International conference on signal propagation and computer technology (iCSPCT)
14. Haines TSF, Xiang T (2012) Background subtraction with dirichlet processes. In: 12th European conference on computer vision
15. Culibrk D, Marques O, Socek D, Kalv H, Furht B (2007) Neural network approach to background modeling for video object segmentation. IEEE Trans Neural Networks
16. Belhani H, Guezouli L (2016) Automatic detection of moving objects in a video surveillance. In: Global Summit on Computer and Information Technology (GSCIT)

6D Pose Estimation of 3D Objects in Scenes with Mutual Similarities and Occlusions

Tiandi Chen and Lifeng Sun

Abstract Estimation of six degrees of freedom (6DoF) attitude of rigid bodies is an essential issue in such fields as robotics and virtual reality. This paper proposed a method that could accurately estimate 6DoF attitude of known rigid bodies with RGB and RGB-D input image data. Additionally, the method could also work well in the scenes in which objects exhibit mutual similarities and occlusion. As one of the contributions made by this paper, a modularized assembly line method was proposed, which integrated deep learning and multi-view geometry method. At first, a neural network for instance segmentation was used to identify the general locations of known objects in the images and give the bounding boxes and masks. Then 6DoF attitude was estimated roughly according to the local features of RGB-D images and templates. Finally, purely geometric method was used to refine the estimation. Another contribution of this paper was the correction of misclassification with the help of some information reserved in the process of training the network. The proposed method achieves a superior performance on a challenging public dataset.

1 Introduction

Estimation of 6DoF attitude of 3D objects has long been an issue arousing study interests in the field of computer vision. This problem is also highly relevant to robotics and virtual reality. With the development of the fields and the improvement of calculation ability, this problem has been gradually developing from simple 2D image detection to complete 6D altitude estimation of objects, including 3DoF location and 3DoF rotation direction. Methods in early period mainly calculated 3D altitude by calculating the feature points of objects according to their surface textures and carrying out matching [1]. The past decade has seen the rapid development of depth

T. Chen · L. Sun (✉)
Department of Computer Science and Technology, Tsinghua University, Beijing, China
e-mail: sunlf@mail.tsinghua.edu.cn

T. Chen
e-mail: chentd11@gmail.com

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_93

camera technique, and post-estimation of texture-less objects becomes more easily. The present studies mainly focus on altitude estimation in difficult scenes with heavy occlusion and other clutter. The solving of this problem will improve the robustness of object recognition algorithm.

For texture-less objects, template-based methods, such as Point-Pair Features and LINEMOD were once popular. The core thought is to obtain the CAD model of a known object and generate the point cloud templates of the object presenting different altitudes. The templates are one-to-one corresponding to poses. Then, local features of point clouds are extracted from the input scenes and compared with the existing templates. Then, ICP algorithm is used for refining the pose, with the final results being output.

The latest study results are already able to address the pose estimation problem of most scenes. Reference [2] improved the point-pair matching method used before and delivered encouraging pose estimation results of rigid bodies in the case of RGB-D input. Reference [3] introduced a new algorithm architecture, which got great recognition accuracy in LINEMOD datasets. The latest dataset T-LESS [4] contained more similar low texture objects and more serious blocking, making the problem more challenging.

There are some similarities between the work in this paper and the work before. Reference [3, 5, 6] all used image data to initialize object coordinate system, while this paper adopted depth data. In addition, the masks of all objects in this paper were known after the first phase, which offered a great advantage.

We summarize our contributions as follows:

- The test result of public datasets showed that our method achieved a good performance, which is much better than the previous methods.
- We proposed a modularized assembly line system, in which modules were not coupled and could work independently, making it possible to improve individual parts.
- We implement a tool to generate new scenes and render it as point cloud data, which generate training and test data randomly and automatically.
- Taking the advantage of this tool, we developed a method to quantitatively measure the similarity between two categories of 3D object, which helps correct the classification error.

The remainder of the paper is organized as follows. Relevant work in this field was introduced in the second part of this paper, detailed introduction to our method was given in the third part, our experiment results were provided in the fourth part, and a summary was made in the fifth part.

2 Related Work

Traditional 2D object detection methods are mainly to match sparse features [1] or templates [7]. Between the two methods, template-based method can better process cases involving low texture objects. Generally speaking, data collected by cameras have small projected areas and texture cannot provide enough feature points. Thus, template-based method usually performs better than feature-based methods. A major drawback of the method completely based on templates is the sensitivity to the occlusion which would reduce the accuracy of recognition and pose estimation.

With the launch of Kinect [8] and other consumer depth transducers, template-based methods can be used more easily. LINEMOD template-based method [9] is a good template-based method. This method combines the features of the normal of point cloud and the normal of RGB image boundary, and generates a descriptor based on the two features, with comparison being made with all template descriptors in the template library. Later, some work optimized the speed of template-based method [10]. Other voting methods have also delivered fine effects, especially point-pair features [11, 2] was also based on point-pair features and made some improvements for the blocking problem that template methods were difficult to address.

The development of deep learning method also leads to breakthroughs in the field of computer vision recognition. Learning-based method is widely applied in the 6D altitude estimation of rigid bodies as well. Reference [12] used neural network to conduct direct regression on pose of objects. But a method completely relying on learning cannot produce highly accurate altitude estimation results. Reference [13] proposed to optimize results by classifying views of discrete objects but did not deliver any test result of the method in challenging public datasets. It seems difficult for pure learning method to achieve a high accuracy of pose estimation. Learning method was only used to figure out the rough locations and masks of the objects rather than directly obtaining 6D pose in an end-to-end network.

Another way to solve the problem was to combine voting and learning together. Reference [14] adopted 6D voting on each image block of RGB-D images using Hough Forest. Reference [15] extracted features with auto-encoder, calculated one feature for each altitude, carried out feature matching on the input RGB-D images, and found out the optimal matches by using the voting method. Random forest, as an improvement of the voting method, can work out the correspondence between images and object coordinates as well. Reference [16] used random forest to match image blocks to the 3D coordinate system and then made corrections by RANSAC.

Reference [17] BB8 was a method proposed recently, which took RGB images as the only input and conducted altitude estimation only based on sparse fixed points of several objects. The pose was estimated by finding the exact points having highly obvious features on object surface. But our experiment results showed that the accuracy could be improved when point cloud information relied on more if there was point cloud data.

All in all, it's difficult to estimate 6D pose of objects in a scene with occlusion. And there were few previous studies that could solve the problem of pose estimation of objects with mutual similarity. However, our method achieved greater accuracy in these cases when RGB-D data input was used.

3 Proposed Method

The framework of our method is shown as follows. First, we used a network (Multi-task Network Cascade, [18]) to conduct instance segmentation for the input images, then found several most similar categories for each object and calculated the similarities of all templates, and weighted the similarities according to the classification error rate obtained in the training of the last step. We choose several most similar templates, which correspond to pose hypothesis. At last, we verify and optimize the pose hypothesis obtained from the template matching of the last step, with the refined results being output (Fig. 1).

The core of the method proposed in this paper was to decompose the problem into three cascade and low-coupling tasks. First, the general location of the object was given according to a detection conducted on 2D image. Then the general pose was provided through comparison between local features of the point cloud and the template point cloud. After that, 6D pose was got through optimization using a pure geometric method. In the first step, the existing instance segmentation method was used to train a deep neural network, as well as classify objects and give the boundary boxes, with the masks of objects being provided at the same time. In the meanwhile, during the training in the first step, some training data and testing data that contained the blocking of objects were generated randomly and automatically, with the error rate of classification being worked out. To some degree, the error rate of classification indicated the similarity level of objects. Based on this, an algorithm could be designed to search among templates of similar object categories in the subsequent template matching process, thus reducing the error rate. In the second step, LINEMOD was used to search among the point cloud templates for the most similar templates of the objects already having masks and falling into some categories. Then, a rough pose estimation value was obtained. In the third step, verifications and refine were made.

In summary, the method in this paper combined the learning method and the traditional geometric method together, which was an advantage. Our goal was to obtain accurate pose estimation, and the mathematic model of the problem is known if the category of the object was determined, so we believe that end-to-end learning

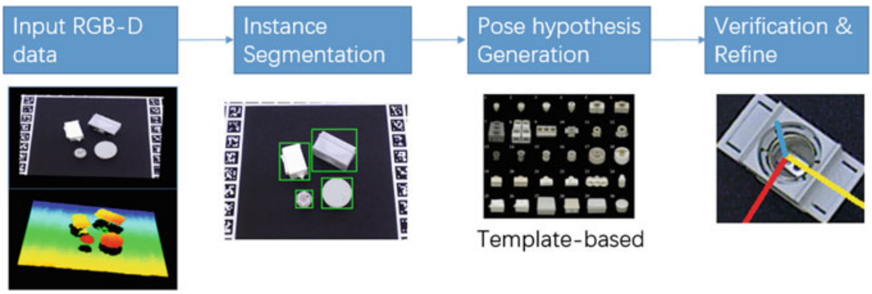


Fig. 1 Architecture of our system

would compromise accuracy. But it was hard for traditional methods to perform a recognition task very well, while the learning method had advantages in this issue. The initial recognition in step one gave the categories and masks of objects, which both reduced the matching times in the template matching stage and mitigated the impacts of noise because of the existence of determined masks. In the subsequent altitude optimization, if the error was found to be large, the smallest error would be found among several alternative results provided in step two and be output, which reduced the calculation work. In this way, both the accuracy and the calculation efficiency were guaranteed.

3.1 Instance Segmentation and Similarity Calculation Between Categories

The main purpose of this step was to find the classification results and masks of all known objects in a scene. The scene was segmented with Multi-task Network Cascade (MNC). If there were K known objects, with an input image being given, the output of the network would contain the masks of all K instances and their categories.

To handle the problem of misclassification, we automatically and randomly generated many scenes with some object was blocked and render them to point cloud data. Then we input the patch of object RGB image to the classifier, and the classification results of which were summarized. The frequency that object i was classified into j was recorded as C_{ij}. $i \neq j$ meant incorrect classification. If the C_{ij} is large, we may need to verify the category of this instance, even calculating the ICP error in the last step to determine which category it belongs to.

3.2 Generating Pose Hypothesis

Here we chose Linemod algorithm to find out the most similar template of the actual pose. Generating templates was a key step in the algorithm. Since 3D CAD models of some objects were not easy to be obtained in actual situations, we developed a method of generating templates of objects having different poses based on the collected RGB-D data. As for the view points, we selected all vertexes of the regular icosahedron on the sphere, as well as the midpoints of all edges of the regular icosahedron. At each view point, we could create a template by rotating through a certain angle for each time. In the experiment, we created a template by moving through every 5°. As a result, $42 \times 72 = 3024$ templates were obtained in total.

After the templates were created, an altitude assumption could be obtained through comparison among similarities between objects in the scenario and templates. Linemod algorithm gave scores to the similarities between each template and the input RGB-D patch of each instance. Generally speaking, a template having the highest score could be taken as the most similar one, with its corresponding pose being taken as the initial value in the following optimization step. However, since there might be incorrect classification, our strategy was to detect all templates in all categories of $C_{ij} > 0.1$, take several templates with the highest similarity scores given by Linemod, and make further judgments at the time of refining the pose in the next step.

3.3 Pose Refine

For an object that was detected in the scene, its mask could help to locate the corresponding part in the point cloud. Those pose hypothesis obtained in the last step were recorded. Iterative Closest Point (ICP) algorithm was performed on the corresponding point cloud Po of each pose assumption, and the result with the smallest matching error was taken as the final output result.

4 Experiments

To test the performance of our system, we conducted a series of experiments. In this section, we first introduce the public dataset we conduct our test on and then show our results. Finally, we show the runtime performance of the system.

4.1 Test Dataset

T-LESS [4] is a new public dataset for testing system performance for 6D pose estimation of objects. This dataset contains 30 types of test objects, most of which are industrial parts without obvious texture. Most of these parts have a single color, so color-based methods cannot work well. The challenge of this data set is that objects are symmetric and mutually similar in shape and size. The test image has 20 different complex scenes, varying from a simple scene composed of several isolated objects to a large number of different objects and other unrelated objects with occlusion.

4.2 Performance

4.2.1 Accuracy of Recognition

We test all 20 scenes of the dataset, and the output of the recognition phase is as follows (Fig. 2):

Because there are many indistinguishable obstructions in some scenes, we just evaluate the instances which have a visible surface area of more than 10% of the total surface area. To figure out the accuracy of recognition, we calculate intersection over union (IoU) of our result and ground truth bounding box, and if the IoU is greater than a threshold (we set the threshold to 0.5), we apply this result as a correct recognition. The accuracy of recognition is as follows (Table 1).

We also figure out the accuracy of recognition of objects belonging to different categories, and we compare our results with BB8 [17]. It should be noted that BB8 only uses RGB data and our method needs depth data (Table 2).

4.2.2 Precision of Pose Estimation

We just sum up 3D location error of instances which is recognized correctly. The average location errors of each scene are as follows (Table 3).

This result achieves a high precision and is not related to the complexity of scenes. So It seems that the bottleneck of the performance is object detection and recognition.

Fig. 2 Recognition result of the instance segmentation network

Table 1 Accuracy of recognition in each scenes

Scene no.	Total accuracy (%)	Scene no.	Total accuracy (%)
1	96.6	11	67.0
2	98.6	12	62.9
3	93.5	13	72.5
4	94.0	14	85.2
5	89.3	15	93.5
6	91.3	16	81.3
7	71.8	17	58.1
8	91.9	18	65.0
9	77.2	19	59.2
10	86.4	20	49.5

Table 2 Compared with BB8

Sense ID: [Obj. IDs]	Accuracy (BB8, RGB only)	Accuracy (Ours, RGB-D)
1: [2, 30]	50.8, 55.4	90.0, 92.0
2: [5, 6]	56.5, 55.6	84.6, 89.3
4: [5, 26, 28]	68.7, 53.3, 40.6	87.3, 76.4, 72.9
5: [1, 10, 27]	39.6, 69.9, 50.1	63.9, 89.3, 88.7
7: [1, 3, 13–18]	42.0, 61.7, 64.5, 40.7, 39.7, 45.7, 50.2, 83.7	77.4, 81.8, 79.5, 71.4, 52.3, 73.0, 83.1, 94.2

Table 3 location error in each scenes. Note that only correct recognition results are applied

Scene no.	Avg. location error (m)	Scene no.	Avg. location error (m)
1	0.0077	11	0.0095
2	0.0059	12	0.0099
3	0.0116	13	0.0173
4	0.0112	14	0.0210
5	0.0085	15	0.0114
6	0.0128	16	0.0106
7	0.0288	17	0.0094
8	0.0167	18	0.0139
9	0.0093	19	0.0127
10	*0.0120*	20	0.0242

Table 4 Time spent of each module

Module	Avg. time for one input scene (ms)
Instance segmentation	11
Pose hypothesis generation	130
Pose refine	6

4.2.3 Runtime Benchmark

At present, the instance segmentation module is implemented by GPU, and other two modules are implemented by CPU with multithread. A computer with an Intel i5-4210H CPU, an NVIDIA GeForce GTX1060 GPU, and an 8 GB memory is used to test. The time spent on every module of the system is as follows (Table 4).

It should be noted that the time required for the second phase is linear with the number of template matches required. So reducing the number of generated templates will speed up the system. If the objects exhibit less mutual similarity, the time spent will also reduce.

5 Conclusion

We proposed a modular system for 6D pose estimation of known objects. Combining deep learning methods and traditional geometry methods, the system achieves a high accuracy of recognition and high precision of pose estimation. And with the help of quantitatively measure the similarity between two categories of 3D object, the system is able to correct the misclassification, which enhances the robustness of the system. Additionally, each module of this system is independent, which makes it easy to modify and improve. From the experiment results, it seems that the bottleneck of performance is still object detection and recognition. The future work will construct a more reasonable neural network for object detection, and optimize the implement for real-time tasks.

References

1. Lowe DG (2001) Local feature view clustering for 3D object recognition. In: CVPR
2. Hinterstoisser S, Lepetit V, Rajkumar N, Konolige K (2016) Going further with point pair features. In: ECCV
3. Michel F, Kirillov A, Brachmann E, Krull A, Gumhold S, Savchynskyy B, Rother C (2017) Global hypothesis generation for 6D object pose estimation. In: CVPR
4. Hodan T et al (2017) T-LESS: An RGB-D dataset for 6D pose estimation of texture-less objects. In: 2017 IEEE Winter Conference on Applications of Computer Vision (WACV), IEEE
5. Brachmann E, Krull A, Michel F, Gumhold S, Shotton J, Rother C (2014) Learning 6D object pose estimation using 3D object coordinates. In: ECCV

6. Brachmann E, Michel F, Krull A, Yang MY, Gumhold S, Rother C (2016) Uncertainty-driven 6D pose estimation of objects and scenes from a single RGB image. In: CVPR
7. Huttenlocher D, Klanderman G, Rucklidge W (1993) Comparing images using the Hausdorff distance. IEEE Trans PAMI
8. WA MCR Kinect for Xbox 360
9. Hinterstoisser S, Holzer S, Cagniart C, Ilic S, Konolige K, Navab N, Lepetit V (2011) Multimodal templates for real-time detection of texture-less objects in heavily cluttered scenes. In: ICCV
10. Rios-Cabrera R, Tuytelaars T (2013) Discriminatively trained templates for 3D object detection: a real time scalable approach. In: ICCV
11. Drost B, Ulrich M, Navab N, Ilic S (2010) Model globally, match locally: Efficient and robust 3D object recognition. In: CVPR
12. Kendall A, Grimes M, Cipolla R (2015) PoseNet: a convolutional network for real-time 6-DoF camera relocalization. In: ICCV
13. Kehl W, Manhardt F, Tombari F, Ilic S, Navab N (2017) SSD6D: making RGB-based 3D detection and 6D pose estimation great again. In: ICCV
14. Tejani A, Tang D, Kouskouridas R, Kim T-K (2014) Latent Class Hough forests for 3D object detection and pose estimation. In: ECCV
15. Kehl W, Milletari F, Tombari F, Ilic S, Navab N (2016) Deep learning of local RGB-D patches for 3D object detection and 6D pose estimation. In: ECCV
16. Brachmann E, Krull A, Michel F, Gumhold S, Shotton J, Rother C (2014) Learning 6D object pose estimation using 3D object coordinates. In: ECCV
17. Rad M, Lepetit V (2017) BB8: A scalable, accurate, robust to partial occlusion method for predicting the 3D poses of challenging objects without using depth. In: ICCV. 1
18. Dai J, He K, Sun J (2016) Instance-aware semantic segmentation via multi-task network cascades. In: CVPR

A Simple and Enhanced Low-Light Image Enhancement Process Using Effective Illumination Mapping Approach

Vallabhuni Vijay, V. Siva Nagaraju, M. Sai Greeshma, B. Revanth Reddy, U. Suresh Kumar and C. Surekha

Abstract When an image is captured in low-light, it gets the low visibility. To overcome the low visibility of the image, some operations are to be performed. But in this paper, image enhancement is introduced using illumination mapping. First, R, G, B maximum values in each pixel of the considered image are to be calculated and then convert it into a grey scale image by applying the formulae. Some filters are used to remove the noise, the choice of filter depends on the type of noise, and then the image is preprocessed. The logarithmic transformation helps to increase the brightness and contrast of the image with a certain amount. Earlier there were some methods to enhance the low-light image, but illumination map existence is chosen. In this illumination, the image will be enhanced with the good quality and efficiency. The illumination technique will be the more efficient and more quality. The illumination corrects the R, G, B values to get the desired image, then Gamma Correction is applied. The Gamma Correction is a non-linear power transform, it helps to increase or decrease the brightness of the desired image when a low value of gamma is taken, the brightness will be increased and when a high value of gamma is taken, and the brightness will be decreased. The proposed system is implemented using MATLAB software. When different types of images are applied, different contrast and brightness levels that depend on the type of image are observed.

1 Introduction

An image captured in low-light conditions is to be enhanced for a better visibility. Image enhancement plays a vital role in improving the digital image quality. Histogram equalization (HE) in [1] is widely used to adjust the image intensities and to enhance the contrast of the image in a variety of applications due to its simple

V. Vijay (✉) · V. Siva Nagaraju · M. Sai Greeshma · B. Revanth Reddy · U. Suresh Kumar
C. Surekha
Department of Electronics and Communication Engineering, Institute of Aeronautical Engineering, Dundigal, Hyderabad 500043, India
e-mail: v.vijay@iare.ac.in

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_94

function and its effectiveness. Contextual and variational contrast enhancement in [2, 3] proposes an algorithm that enhances the contrast of input image using interpixel contextual information. It maps the elements of one histogram to elements of another histogram diagonally. Layered difference representation is used to enhance the image contrast by mapping the input grey levels to the output grey levels. The grey-level difference between the adjacent pixels is amplified. Naturalness preserved enhancement algorithm preserves the naturalness of the image by using three methods, namely, lightness-order-error measure, bright pass filter and bi-log transformation. These are used for preserving naturalness and also enhances the image. Applications for the above-mentioned methods are that these are applied for thresholding, normalization of magnetic resonance imaging (MRI) images. For comparing two or more images on a particular basis, histograms of the images are to be normalized.

The major drawbacks of existing methods are an increase in the contrast of noise by decreasing the quality of the image. The results are computationally intensive. When we increase the brightness of the image, the brighter particles in the image will get brighter. One major drawback observed is with rapid change in technology, techniques that are used in real-time applications always keeps on changing less than a decade, i.e., digital image processing's real-time products can only survive in the market for a maximum of 3–5 years only and in this span other technologies arises. The initial cost of the technique is high depending on the type of system.

Proposed work mainly describes input image, that is, the low-light captured image is chosen and is to be enhanced to obtain the desired modified image as said in [4]. The file path of the image to be enhanced is considered and is preprocessed so that noise can be reduced easily. Noise present in the image causes blurriness. These noise or distortions are filtered using preprocessing step. The uneven illumination caused by sensors, default, non-uniform illumination of the scene or orientation of objects can be rectified using illumination correlation. Prospective correlation, retrospective correlation, and other using low pass filtering as from [5] is also done using illumination correction [6]. In illumination correction brightness and contrast can be applied to the image whereever required non-linear distribution is done. The greyscale levels between black and white colours are enhanced using gamma correction [7]. Gamma correction also enables us to compare grey levels between the adjacent pixels and stop adjust whichever colour is required. Pixel intensity enhancement has to be carefully adjusting brightness or contrast levels to larger extend may cause irregularities in the image. These pixels enhancement can be applied to both grey and colour scale image. Illumination map estimation is done using RGB channels individually to the input image and then is converted to grey scale image to check the improvements in the image. Enhancement is done until the desired output is obtained. Advantages and applications are also explained in this paper.

This paper is well organized as it clearly states step to step process of this project. Image enhancement is the major part as it is a basic need when any image is considered. The drawbacks in previous methods and the possibility to overcome them are discussed in the above paragraphs. Reduction of noise as explained in [8, 9], better version of HE and other steps has been taken. In the above-mentioned theory, it clearly states that what the steps are done from the beginning input image to obtaining the desired output. Section 2 describes how the input image is taken, reduction of noise, deblurring the blurred image, etc., are explained individually. Adjusting the brightness and contrast of the image is done accordingly. Applying different filters individually for the considered input image. Gamma correction enhances the pixel intensity. Gaussian and median filters [10] are used for noise reduction. Recovering the desired image from noise affected or blurred features can also be achieved [11]–17]. In continuation with Sects. 2, 3 deals with the simulation results and comparison table of the input and output images.

2 Design Methodology and Observation

The design methodology briefly describes how image enhancement, preprocessing, transformation, illumination correction and gamma correction are performed on the image which is taken under low-light conditions. The enhanced image is obtained by using the below-explained methods.

2.1 Image Enhancement

Image enhancement is the process by which the digital images can be adjusted for better quality and for further use. Here the noise can be reduced, the images are sharpened and brightened.

The various methods involved in image enhancement are morphological operations, HE for removal of noise wiener filter is used, linear contrast adjustment, median filtering, unsharp mask filtering and decorrelation stretch [1].

Advantages of image enhancement are manipulating the pixel values in the image and it is easy for visual interpretation [2].

The flow chart is given in Fig. 1 describes each and every step of the techniques which are used for enhancement of the image. The algorithm shown below gives the information about how an image is taken and processed in a step by step manner. The first is to reduce the noise and enhancing the image by using some of the transforms and in the final step the image contrast levels are increased by using gamma correction.

Algorithm: Image Enhancement

Input: The input image is given by $F(x, y)$.

Initialization:

Update the value of $R(x, y)$ using equation (1).

Update the value of $g(x, y)$ using equation (2).

Update the value of $T_w(f)$ and $T_b(f)$ using $g(x, y)$.

Output: The solution is s.

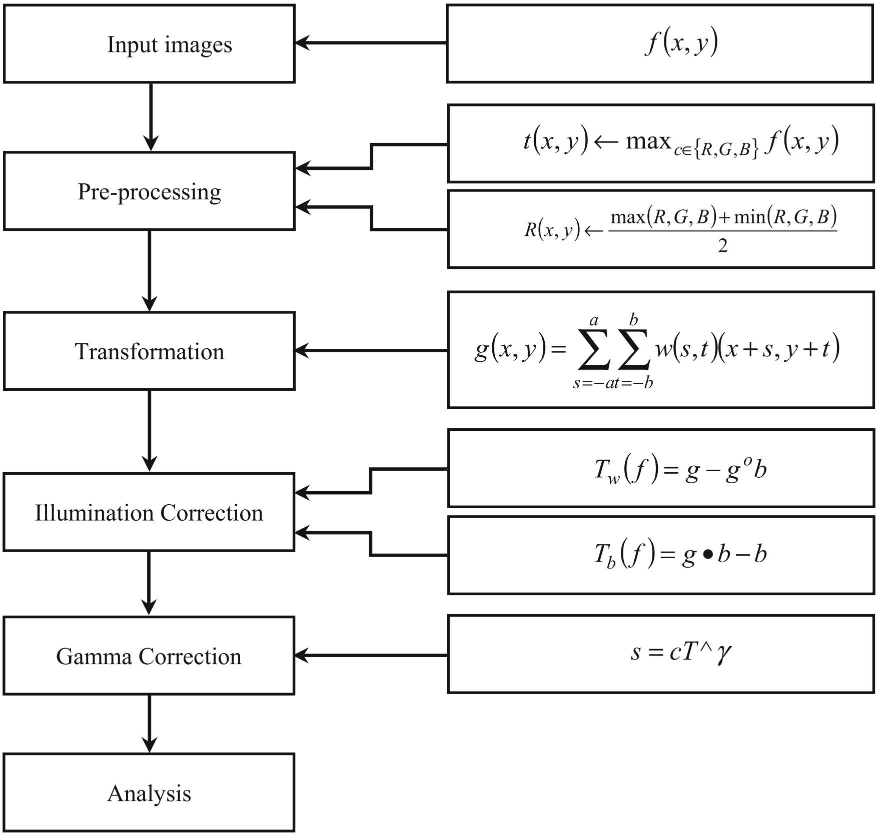


Fig. 1 Flow chart

2.2 Preprocessing

Preprocessing is a technique used to suppress the noise and distortion which are present in the image and it enhances the information in the image. The median filter is used in reducing the noise. It is a non-linear digital filtering technique. Reduction of noise is a major solution in preprocessing step for improvement of further use. Figure 2b shows how noise is reduced. By using a median filter, it preserves edges when noise is reduced. The main advantage of preprocessing is to reduce the noise [4].

$$t(x, y) \leftarrow \max_{c\in\{R,G,B\}} f(x, y) \tag{1}$$

$$R(x, y) \leftarrow \frac{\max(R, G, B) + \min(R, G, B)}{2} \tag{2}$$

(a) Before filtering (b) After filtering

Fig. 2 Removal of noise using a median filter

(a) Before transformation (b) After transformation

Fig. 3 Image after transformation

The two Eqs. (1) and (2) represent the input image and the image converted into a grey image. Here f (x,y) represents the input image and R (x,y) is the grey image.

2.3 *Transformation*

Transformation is a technique used for operations in the image. First, the image is transformed into spatial domain then it removes the noise and again convert it into the frequency domain. The image before and after transformation is shown in Fig. 3a and b. The mathematical representation of the spatial domain is given in Eq. (3).

$$g(x, y) = \sum_{s=-a}^{a} \sum_{t=-b}^{b} w(s, t)(x + s, y + t) \tag{3}$$

where $w(s, t)$ is the image which is a preprocessed image, $R(x, y)$ is the grey image and $g(x, y)$ is the transformed image.

Advantages of transformation are noise reduction. In the frequency domain, fast convolution is carried out [5].

2.4 Illumination Correction

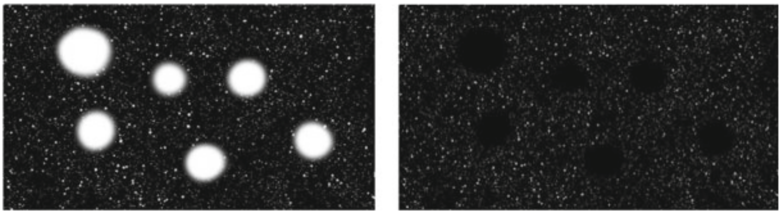
It is used to adjust the brightness and contrast level of the image. Illumination correction is done by two transforms, namely, top-hat and bot-hat transform. Top-hat transform is used in digital image processing and mathematical morphology. It is an operation used for the extraction of small elements that are given in Fig. 4.

It is also used for extracting the details from the given image. There are two types of top-hat transforms, namely, (1) white top-hat transform, (2) black top-hat transform. The white top-hat transform is known as the difference between the input image and the structuring element by opening it. White top-hat transform is represented by $T_w(f)$. The black top-hat transform is known as the difference between the closing of structuring element and input image and it is represented by $T_b(f)$. In various applications, top-hat transform is used [7].

$$T_w(f) = g - g^o b \tag{4}$$

$$T_b(f) = g \bullet b - b \tag{5}$$

The above two Eqs. (4) and (5) are top-hat and bot-hat transforms, where g is transformed image and b is the structuring element.



(a) Before transform (b) After transform

Fig. 4 Illumination correction

(a) Before applying gamma (b) After applying gamma

Fig. 5 Gamma correction

2.5 Gamma Correction

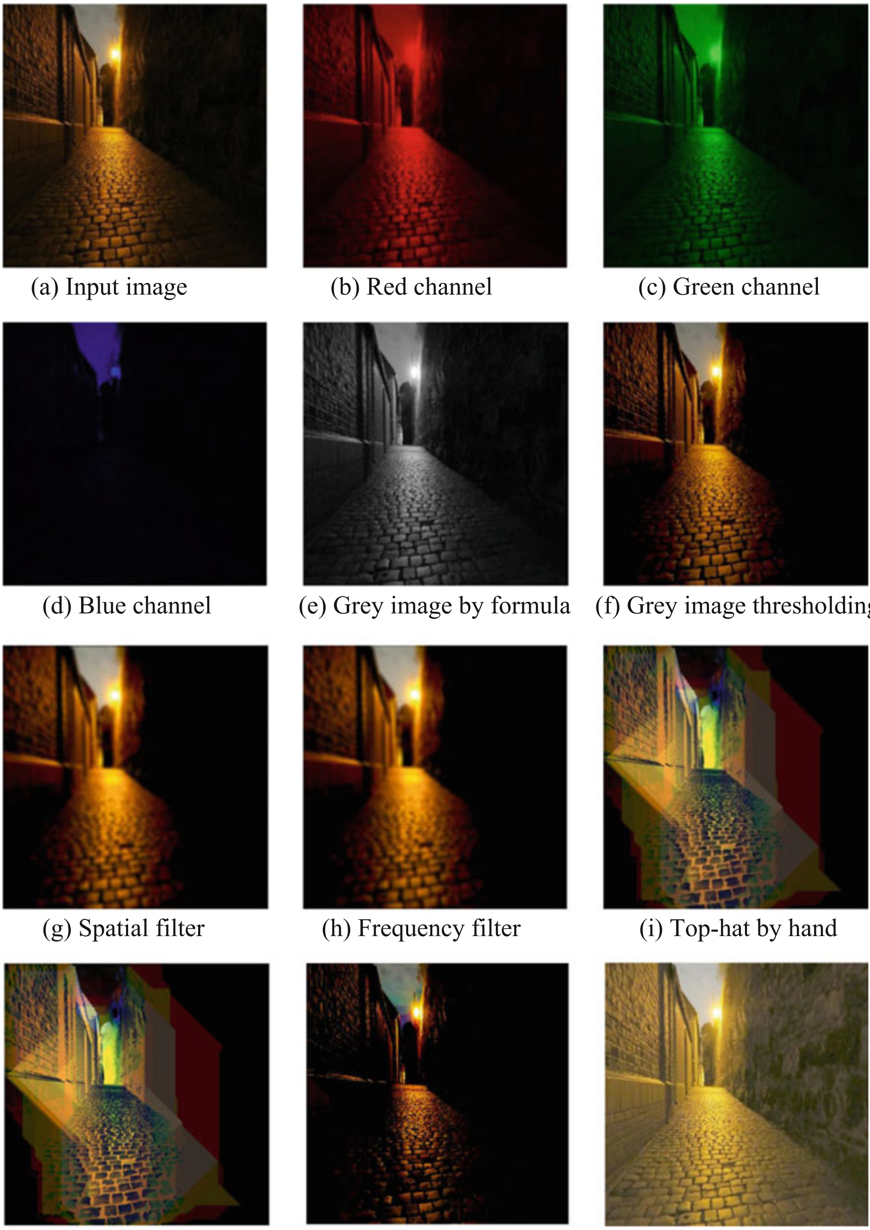
Gamma correction is also known as power law transformation. The nth power and nth root transformations are two types of power law transformations. These transformations are given by the expression presented in Eq. (6).

$$s = cT \wedge \gamma \tag{6}$$

This γ symbol is called gamma, by reflecting this symbol name, it is also known as gamma transformation. By varying the γ values the image can be enhanced that is shown in Fig. 5. Here c is a colour component value ranging from 0 raised to some power γ and T is the image. Different devices or monitors have different display settings, by this, the intensity also changes. The advantage of gamma is that the image can be displayed on CRT and LCD [8].

3 Simulation Results of the Proposed Design

Figure. 6a-l represents each and every method in the proposed model about how the image is taken and enhanced. First, the input image is converted into RGB channels and then converted into a grey image. The grey image is then preprocessed using a median filter and the image is transformed into frequency domain [5]. The image in [7] frequency domain is enhanced by changing the contrast levels using top-hat and bot-hat transform and next to the image is corrected by using gamma correction; in this, the image is enhanced by changing the gamma values. Thus, the output image is obtained shown in Fig. 6(l).



(a) Input image (b) Red channel (c) Green channel

(d) Blue channel (e) Grey image by formula (f) Grey image thresholding

(g) Spatial filter (h) Frequency filter (i) Top-hat by hand

(j) Bot-hat by hand (k) Top-hat minus bot-hat (l) Output image

Fig. 6 Simulation Results

Table 1 Comparison table

Type of image	MSE	SNR
Input image	1107.9	0.6948
Red channel	1223.9	0.8791
Green channel	1242.9	0.8463
Blue channel	1278.3	0.7873
Grey image by formulae	1148.5	1.0286
Grey image threshold	1228.7	0.9052
Spatial filter	1217.4	0.9286
Frequency filter	1215.7	0.9281
Top-hat transform	1211.1	0.9055
Output image	1133.05	1.1090
Input image	1107.9	0.6948

The comparison of different types of images is done by calculating the mean square error (MSE) and signal to noise ratio (SNR). The MSE and SNR of the input image are 1107.9, 0.6948. The MSE and SNR of the output image are 411.90, 1.55, respectively. It is as shown in Table 1.

4 Conclusion

In this paper, a low-light image enhancement is proposed. By decomposing a low-light image into the illumination component, it offers a solution to expand illumination and enhances image details separately. Specifically, the illumination component is processed using median image filter in gradient domain, followed by a non-linear logarithmic transform. Then the illumination component is enhanced by the gamma transform. This leads to enhance the low-light images and effectively reduce the distortions in colour image and also reduces noise and image blurring. Then, the final result is analyzed by the output of the enhanced illumination component. Experimental results show that the enhanced images produced by the proposed method are visually clear and more effective by the performance of the proposed method outperforms the existing methods in terms of both HE and Adaptive histogram equalization (AHE) assessments. Moreover, the proposed algorithm is efficient because the computation complexity is not related to filter size. The proposed method has great potential to implement in a real-time low-light video processing.

References

1. Ghitta O, Ilea DE, Whelan PF (2013) Texture enhanced histogram equalization using TV-L1 ımage decomposition. IEEE Trans Image Process 22(8):3133–3144
2. Celik T, Tjahjadi T (2011) Contextual and Variational Contrast Enhancement. IEEE Trans Image Process 20(12):3431–3441
3. Luo Y, Guan Y-P (2015) Structuralcompensation enhancement method fornonuniform illumination images. Appl Opt 54(10):2929–2938
4. Guo X, Li Y, Ling H (2017) LIME: Low-Light Image Enhancement via Illumination Map Estimation. IEEE Trans Image Process 26(2):982–993
5. Low pass filters, https://www.picosecond.com/objects/AN
6. Yang J, Zhong W, Miao Z (2016) On the Image enhancement histogram processing. In: 3rd ınternational conference on ınformative and cybernetics for computational social systems (ICCSS). Jinzhou, China, pp 252–255
7. Kubinger W, Vincze M, Ayromiou M (1998) The role of gamma correction in colour image processing. In: 9th European signal processing conference (EUSIPCO 1998). Vienna, Austria, pp 1–4
8. Noise reduction filters for image processing, https://www.sciencedirect.com/science/article/pii/S1875389212005494
9. Filters for noise reduction, http://www.radiomuseum.org/forumdata/users/4767/file/Tektronix_VerticalAmplifierCircuits_Part1.pdf
10. Huang T, Yang G, Tang G (2014) A fast two-dimensional median filtering algorithm. IEEE Trans Acoust Speech Signal Process 27(1):13–18
11. Gehler P, Rother C, Kiefel M, Zhang L, Scholkopf B (2011) Recovering intrinsic images with a global sparsity prior on reflectance. In: Neural ınformation processing systems. California, United States, pp 765–773
12. Enhancement methods in image processing, https://in.mathworks.com/discovery/image-enhancement.html
13. Analysis of image enhancement, http://acharya.ac.in/aigs/firstissuepapers/paper7.pdf
14. Image pre-processing, https://www.slideshare.net/ASHI14march/image-pre-processing
15. Image transformations, https://www.tutorialspoint.com/dip/image_transformations.htm
16. Top-hat transform, https://en.wikipedia.org/wiki/Top-hat_transform
17. Image processing algorithms part 6: gamma correction, http://www.dfstudios.co.uk/articles/programming/image-programming-algorithms/image-processing-algorithms-part-6-gamma-correction/

Comparison of Particle Swarm Optimization and Weighted Artificial Bee Colony Techniques in Classification of Dementia Using MRI Images

N. Bharanidharan and Harikumar Rajaguru

Abstract Numerous soft computing techniques are used nowadays to analyze medical images, and diagnosis of disease is computerized. This paper compares the performance of Weighted Artificial Bee Colony and Particle Swarm Optimization in the diagnosis of dementia using MRI images. For analysis, cross-sectional MRI of 235 subjects collected from OASIS is used. By adjusting the weights for both optimization techniques in a proper manner, optimized results can be reached. These techniques classify the cross-sectional image into three categories and give almost equal Goodness Detection Ratio of 78% along with different regression ratios.

1 Introduction

Magnetic Resonance Imaging (MRI) is a popular and robust visualization technique which helps to characterize the internal structures of the human body in a secured way [1]. In the diagnosis of various brain disorders, MRI plays a vital role due to the potential of recording data in a distinguished manner for dissimilar soft tissues. The usefulness of MRI images is appreciably exaggerated when automated, precise classification is possible.

Dementia is an overall term which denotes the group of symptoms which causes gradual declination of memory or thinking/reasoning skills. Other symptoms of dementia include notable difference in personality and emotional regulation. This disease is so common to aged peoples and the people affected with dementia are presumed to grow from 35 million today to 65 million by the year 2030 [2].

In dementia diagnosis, radiologists have to integrate brain MRI scans along with their medical knowledge to know the severity of dementia and to decide the treat-

N. Bharanidharan (✉) · H. Rajaguru
Department of ECE, Bannari Amman Institute of Technology, Sathyamangalam, India
e-mail: bharanidharan@bitsathy.ac.in

H. Rajaguru
e-mail: harikumarr@bitsathy.ac.in

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_95

ment method [3]. But the greater difficulty in MRI-based dementia detection will be large number of scanned images for each patient which causes diagnosing dementia manually a tedious process. Hence, computer-aided dementia diagnosis methods are essential to increase the diagnostic correctness. Numerous algorithms have evolved in the past three decades to solve this problem, but still there is a vacuum for an accurate and computerized technique to help dementia detection.

The paper is organized as follows: Second section presents Particle Swarm Optimization (PSO) algorithm and third section explains the original and weighted Artificial Bee Colony (ABC) technique and fourth section deals with classification of dementia and results are analyzed in the last section.

2 Particle Swarm Optimization Technique

Evolutionary Algorithm (EA) imitates principles of uninhibited evolution to perform optimization process in numerous methods. Replicating the natural character of bird flocks, this population-based swarm intelligence optimization technique was modelled. Group of random particles (solutions) are assigned for PSO initially and then the optimum solution is found iteratively by renovating its position and velocity [4–6]. Every particle will have both social learning and cognitive learning, i.e. they change its present position depending on its distance from the global best and its personal best. Depending on the problem of optimization, assessment of closeness between the particle and global optimum, i.e. fitness function is chosen. Each particle i carries the following data during flight: x_i, ith particle's current position (x-vector), v_i, ith particle's current velocity (v-vector). p_i (p-vector).

The personal best (pbest) position of a particular particle is the finest position which the particle i, visited so far, i.e. a position of particular particle which gave maximum fitness value so far. Consider f as the fitness function, and then the personal best position of ith particle at a time step t is updated as

$$p_i(t+1) = \begin{cases} p_i(t) & if\ f(x_i(t+1)) \geq f(p_i(t)) \\ x_i(t+1) & if\ f(x_i(t+1)) < f(p_i(t)) \end{cases} \tag{1}$$

Information sharing among swarm members is the important characteristics of PSO. This sharing of information helps to calculate the best position that enables social learning. In PSO, most of the times, either ring or star topology will be used to share information among neighbours.

Let gbest be the postion of global best, then

$$gbest \in \{p_0(t), p_1(t), \ldots, p_m(t)\} = \min\{f(p_0(t)), f(p_1(t)), \ldots, f(p_m(t))\} \tag{2}$$

The velocity and position of ith particle are renovated using the following equations:

$$v_i(t+1) = w * v_i(t) + c_1 * r_1 * (p_i(t) - x_i(t) + c_2 * r_2 * (gbest - x_i(t)) \quad (3)$$

$$x_i(t+1) = x_i(t) + v_i(t+1) \quad (4)$$

Here w is inertia weight while c_1 and c_2 are cognitive learning constant and social learning constant, respectively. The parameters r_1 and r_2 are arbitrary numbers with value ranging from 0 to 1 to avoid local minima problem. To reduce the effect of the previous velocity, inertia term is used. A very little inertia term generally pulls exploitation, while a very huge inertia term may pull exploration. The performance of PSO is extremely depending on the parameters $c1$, $c2$ and w. Iteratively, PSO changes velocity and position and repeats until velocity changes are nearing zero or a particular number of maximum iterations have been reached.

3 Artificial Bee Colony

This algorithm categorizes bees in three groups: employed bees, scouts and onlookers. Employed bees will search for food and pass their food information to onlookers. Onlookers select good food source using the information shared by employed bees. If the eminence of the particular food source is not enhanced, then it is abandoned. Every food source (x_i) corresponds to a feasible solution of the optimization problem and fitness value ($f(x_i)$) is used to know the nectar amount of a food source [7, 8]. The count of employed bees is chosen equal to the number of food sources. Based on the optimization problem, the food sources are initialized.

The nectar information is evaluated by onlooker bees and the food source with higher quality will be chosen by onlookers. The food quality can be calculated using the following equation:

$$p_i = \frac{f(x_i)}{\sum_{i=1}^{n} f(x_i)} \quad (5)$$

3.1 Methodology for ABC

When a food source x_i is chosen using any one selection method, update on the solution in her memory will be done based on the equation:

$$x_i' = x_i + \alpha(x_i - x_k) \quad (6)$$

where x_i' is a new feasible solution produced from its previous solution x_i and the randomly selected neighbouring solution x_k; α is a random number between [− 1, 1].

3.2 Weighted ABC

In the original ABC algorithm, random number (α) is used to eradicate local minima problem by introducing some randomness. But there should be controlling weight parameter for this randomness in a classification problem. In proposed weighted ABC, position is updated using equation given as

$$x_i(t+1) = x_i(t) + a_1 * r_1 * (x_i(t) - x_k(t)) + b_1 * r_2 * (x_i(t-1) - x_k(t-1)) \quad (7)$$

where t denotes the iteration step, r_1 and r_2 are random numbers in the range [0, 1], k denotes the neighbour chosen randomly, a_1 and b_1 are the control weights.

4 Classification of Dementia

OASIS furnishes brain MRI images related to dementia comprising cross-sectional brain images of 416 patients with age ranging between 18 and 96. Clinical Dementia Rating (CDR) of 235 patients is accessible, out of 416 patients. Each subject under evaluation will be categorized as Non-Dementia (ND), Very Mild Dementia (VMD) or Mild Dementia (MD). Targets are fixed for each class during the training and testing phase as shown in Table 1.

During both training and testing, MRI brain image of every patient is split into 16 sub-images and statistical features like mean, variance, skewness, kurtosis features are obtained (Fig. 1).

4.1 PSO Based Classification

For each image 64 normalized features are initialized as 64 particles (solutions). Velocity of all solutions is arbitrarily assigned as 0.5 and maximum number of iterations are used as stopping criterion. PSO nears the target precisely with more number of iterations and so optimal value selection of maximum number of iterations is required. The optimal values for w, c_1 and c_2 are selected while training and kept as constant during valuation.

Table 1 Data set and its target

	ND	VMD	MD
Training	63	35	15
Testing	62	35	15
Target	0.1	0.6	0.9

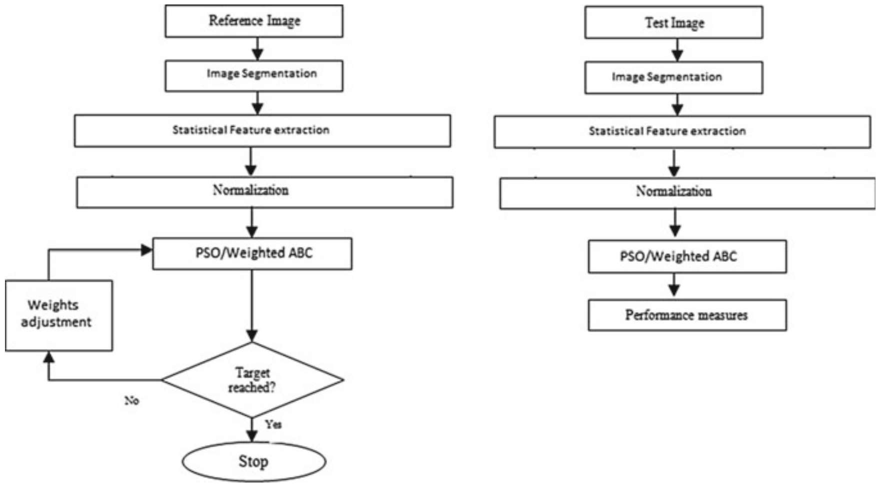


Fig. 1 Flow chart of training and testing the dementia classifier based on PSO/Weighted ABC

4.2 Weighted ABC Based Classification

For each image, 64 normalized features are initialized as 64 food sources. Similar to PSO, maximum number of iterations are used as a stopping criterion. Roulette wheel based selection approach is used for selecting the food source with high nectar quantity. The optimal values for a_1 and b_1 are chosen while training and kept as constant during valuation (Fig. 2).

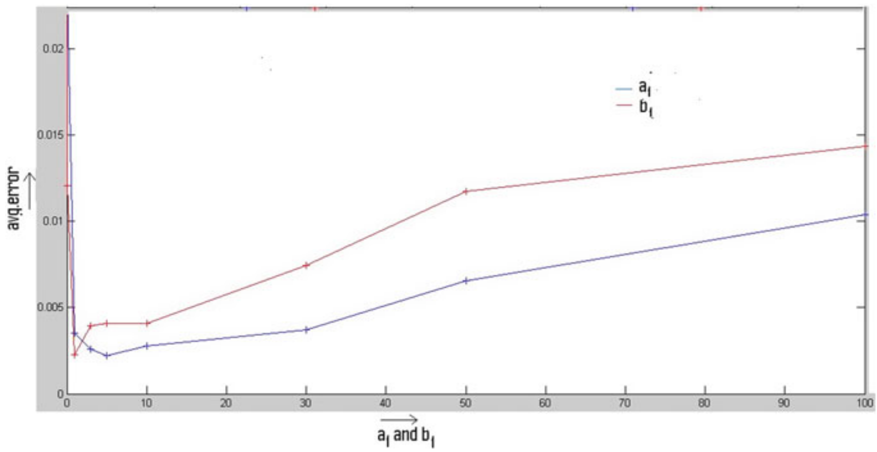


Fig. 2 Average error response of weighted ABC for variation in a_1 and b_1 weights

5 Results and Discussion

To examine the performance of the classifier, Ratios of Regression (R_1 and R_2) and Goodness Detection Ratio (GDR) are determined from the confusion matrix. A classifier is named as good if the GDR crosses 75%.

$$GDR = \frac{PC - MC}{PC + FA} \times 100\% \tag{8}$$

$$R_1 = \frac{MC}{FA} \tag{9}$$

$$R_2 = 1 - \frac{FA}{MC} \tag{10}$$

In the above equations, PC means Perfect Classification, MC means Missed Classification and FA means False Alarm.

From Fig. 3, it is observed that convergence speed of weighted ABC is high compared to the PSO. Hence the computational time of weightedABC will be less due to reduced number of iterations. On the other hand, average error will not be equal to zero for weighted ABC even after 1000 iterations while PSO approaches zero average error with 100 iterations (Table 2).

From the above results, GDR will be equal to 77.57% and Ratios of regression, $R_1 = 0.714$, $R_2 = -0.4$ for PSO-based classifier. For Weighted ABC based classifier, GDR = 77.22%, $R_1 = 2.28$ and $R_2 = 0.562$.

Thus, the overall classification accuracy is same for PSO and Weighted ABC based classifier if optimal weights are chosen. But there is a huge difference in regression ratio between the two types. It is due to high false alarm rate and low missed classification rate of PSO-based classifier; high missed classification rate and low false alarm rate of Weighted ABC based classifier.

Table 2 Confusion matrix of PSO and weighted ABC based classifier

Actual/ predicted class	PSO based classifier			Weighted ABC based classifier		
	ND	VMD	MD	ND	VMD	MD
ND	61	6	0	55	12	0
VMD	4	23	8	3	28	4
MD	0	6	9	0	4	11

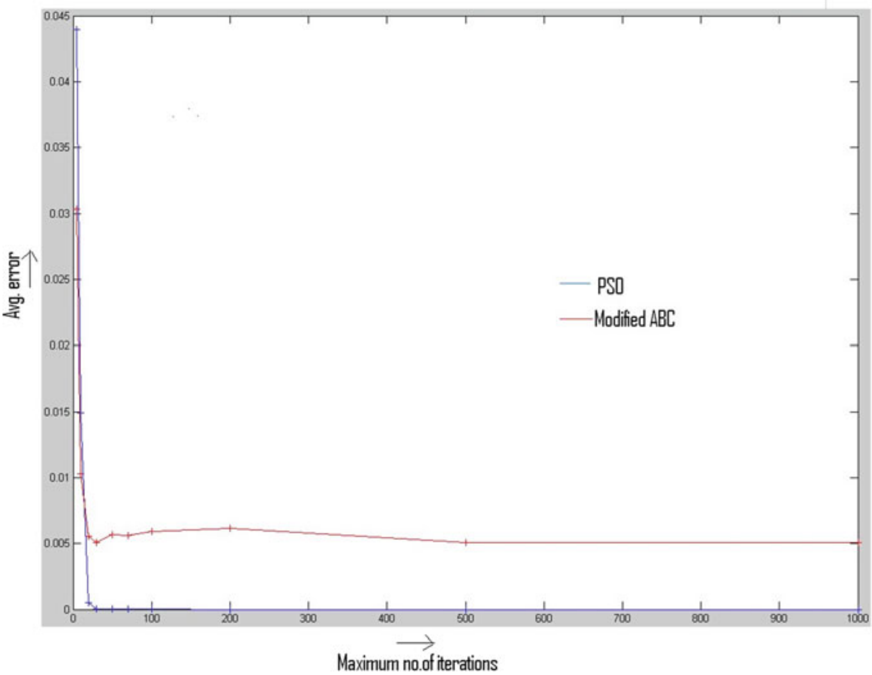


Fig. 3 Comparison of PSO and weighted ABC using average error response for variation in maximum number of iterations

6 Conclusion

The performance of PSO and weighted ABC based classifier in dementia classification is analyzed. GDR can be increased by tuning the values of weights of PSO and weighted ABC optimally. Improper selection of weights may result in low GDR. Performance of weighted ABC and PSO based classification methods can be enhanced taking into consideration other parameters like age, whole brain volume, total intracranial volume, etc. Hybrid classifiers such as ABC-PSO and ACO-PSO will be the direction of future research work.

References

1. Rodriguez AO (2004) Principles of magnetic resonance imaging. Rev Mex Fis 50:272–286
2. Korolev IO (2014) Alzheimer's disease: a clinical and basic science review. Med Stud Res J 04 (2014)
3. Kapse RS, Salankar SS, Babar M (2015) Literature survey on detection of brain tumor from MRI images. IOSR J Electron Commun Eng (IOSR-JECE) 10:80–86

4. Kennedy J, Eberhart R (1995) Particle swarm optimization. IEEE Int Conf Neural Netw Australia 4:1942–1948
5. Mohsen F, Hadhoud M, Mostafa K, Amin K (2012) A new image segmentation method based on particle swarm optimization. Int Arab J Inf Technol 9
6. Omran M, Engelbrecht A, Salman AA (2005) Particle swarm optimization method for image clustering. Int J Pattern Recogn Artif Intell. https://doi.org/10.1142/S0218001405004083
7. Chun-Feng W, Kui L, Pei-Ping S (2014) Hybrid artificial bee colony algorithm and particle swarm search for global optimization. Math Prob Eng Article ID 832949
8. Cao L, Xue D (2015) Research on modified artificial bee colony clustering algorithm. In: International conference on network and information systems for computers. https://doi.org/10.1109/ICNISC.2015.62

Automatic Segmentation of Malaria Affected Erythrocyte in Thin Blood Films

Komal B. Rode and Sangita D. Bharkad

Abstract In today's world, a highly précised diagnostic method needs to be improved for management of feverish sickness and ensure that medicines are prescribed when necessary. The proposed algorithm is applied to giemsa-stained thin blood films. Using triangle's thresholding technique, the erythrocytes are segmented, HSV color space based feature extraction is applied on the segmented erythrocytes. Features extracted are given to SVM classifier to identify whether the query sample is affected with a parasite or a normal sample. The performance of this algorithm is evaluated on the database collected from CDC website and 20 samples taken manually from a local hospital.

1 Introduction

Malaria is a feverish illness and a main public health issue worldwide. It is evoked by a parasitic protozoan of the genre plasmodium transferred through the bite of infected female Anopheles mosquito. It has an impingement on children less than 5 years of age and pregnant women contributing significantly to maternal deaths, low-birth weights in infants [1]. There are five different malaria species *p. falciparum*, *p. vivax*, *p. malariae*, *p. ovale*, and *p. knowlesi* which are harmful to human. *p. falciparum* is one of the deadly species which if not detected before time may lead to serious health issues. Blood smear image of a normal person consists of as RBCs, WBCs, platelets, and the liquid plasma as main constituents. When a person is suffering from malaria the parasites dwell in RBCs. Figure 1 shows the microscopic malaria blood sample image which shows the ring-like structure of malaria parasites and the normal red blood cells.

K. B. Rode (✉) · S. D. Bharkad
Department of Electronics and Telecommunication, Government College of Engineering, Aurangabad, India
e-mail: komal18oct@gmail.com

S. D. Bharkad
e-mail: sangita.bharkad@gmail.com

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_96

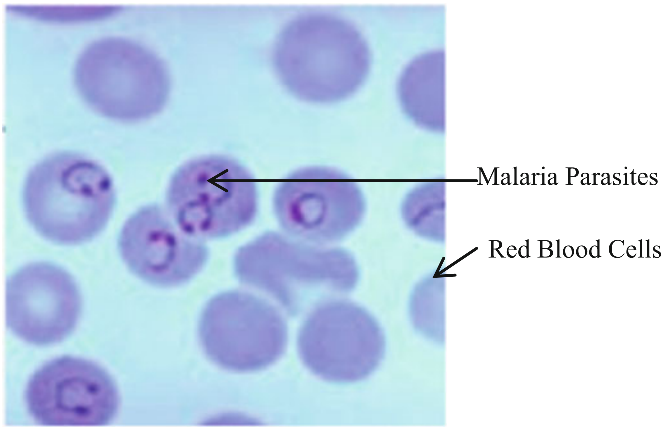


Fig. 1 Microscopic blood sample

Table 1 Pathology methods for malaria test [3]

Sr. no	Name of test	Time required (min)
1	PBS	30
2	QBS	10–15
3	RDT	10–15
4	PCR	45

Traditionally, to validate the presence of parasites different stains such as Giemsa, Leishman, AB are used depending on the environmental conditions. The peripheral smear preparation consists of two approaches thick smear and thin smear. The pathologist uses thick smear for parasitemia estimation and thin smear for species estimation. Staining is a process which highlights the parasites, making it easier for the expertise to identify the presence of parasites in the RBCs. The parasites appear dark (the color depends upon the stain) in color due to the presence of chromatin dots while erythrocytes are light in color (as the nucleus is absent), the background contains plasma lightly colored [2, 3]. The pure standard for malaria parasite estimation is smear preparation which is based on microscopic examination. Different clinical diagnosis methods such as PBS (peripheral blood smear), serological tests, QBS (quantitative buffy coat), and RDT (Rapid diagnostic test) are used for examining the presence of malaria parasites in blood samples. The efficiency of laboratory techniques mainly depends on factors like sensitivity, a number of parasites per microlitre of blood, time required, and cost. Table 1 describes the clinical malaria diagnosis methods and time required to perform the respective tests.

However, clinical diagnostic tests are based on visual inspection by the medical experts. The visual analysis can be hard. Hence, the outcome of diagnostic results on a visual assumption due to human interactions may sometimes be prone to error.

1.1 Motivation

With the advance in technology, many computer-aided techniques have been successfully developed for the reason of facilitating and accelerating the clinical diagnosis test. Today, reducing the time factor is a major concern for malaria diagnosis and making it manual interpretation independent. Proper medication can be given only when the clinical test is a success. New kits available suffer from flaws like temperature tolerance, shows false negative count, and their sensitivity vary considerably [3]. Reducing the flaws in kits available and a new automatic diagnosis tool is a vital factor in finding diseases. By reducing the manual interruption, medical expertise can focus their attention on the patients with a positive test. Therefore, we aim to propose an automatic method which identifies the affected RBC's with maximum accuracy. The key points of our algorithm are the triangle's thresholding method [4] which separates the foreground and background to identify the region of interest.

1.2 Related Work

Many researchers have proposed different image processing related work for parasite identification. Yashasvi et al. [5] used chan-vese segmentation method for boundary extraction of the cells. Hough transform to count the RBCs with k-means clustering for parasite detection. Gloria et al. [6] proposed a method which consists of the preprocessing step for luminance correction. A segmentation technique by inclusion tree representation is used to separate the erythrocytes from the background. Recognition of erythrocytes is done using color pixel classification. Feature extraction using the color histogram, Tamura texture, and saturation histogram is performed. For classification of parasites, radial basis and polynomial kernel function are considered. The amount of parasitemia present is obtained using a trained bank of the classifier. This technique has 99.7% specificity and 94% sensitivity. Boray et al. [7] have given a brief review of the computer vision techniques with partial solutions related to diagnostic techniques for malaria parasite detection. Minh-Tam et al. [8] used Zack's thresholding for extraction of the nucleated portion. All the solid constituents are identified by comparing the malaria image with an empty field image. The size of erythrocytes is obtained. Based on isolated erythrocytes area, the leukocytes and gametocytes are detected and isolated accordingly. Kareem et al. [9] mentioned a method that differentiates gametocytes using modified annular ring ratio. Kareem et al. [10] described a method in which depending on the geometry and color features of the erythrocytes the parasites are identified with good accuracy results. Literature mentioned above mainly focus on identifying the parasites present in the erythrocytes; working in unsupervised conditions are not addressed. Understanding the manual diagnosis method need of algorithm is to reduce the medical expertise intervention and time factor. Hence, with focus on above factors, an automatic algorithm is proposed.

Rest of the paper describes the method implemented which is organized as follows. Section 2 describes the proposed algorithm. Section 3 presents the experimental results of the proposed approach. Finally, the implementation is concluded in Sect. 4.

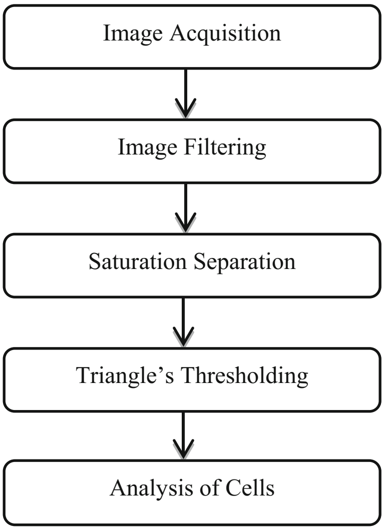
2 Proposed Approach

Figure 2 presents the flow of the proposed work. The approach is a simple algorithm which segments the cells present in each image and classifies whether the cell is infected with malaria parasites or normal RBC.

2.1 *Image Acquisition*

For image acquisition, first step is slide preparation by an expert for manual examination. Giemsa staining technique is used, dilution is made in alkaline buffer with pH 7.2. After staining slide is rinsed under tap water and dried. Slides are prepared to detect the count of parasitemia under a microscope by lab expertise. Images were acquired by placing the high-definition digital camera with an extension to a microscope. For this work, a 12-megapixel resolution camera was used for acquiring the sample images.

Fig. 2 Proposed approach



2.2 Average Filtering

Filtering is a preprocessing method which aims in the suppression of image noise, detect edges, etc. Averaging filters smoothens the image by reducing intensity variation in neighbor pixels. It reduces noise, suppresses irrelevant detail in the image, and blurs the image which makes the transition smooth from one color to another. It works replacing each pixel to pixel by its average value with a 3 by 3 filter mask. The mask gives equal weights to all the pixels. Average filtering is chosen since it blurs the image noise components introduced, i.e., artifacts which are smaller in size. These artifacts should be blended because they increase chances of false prediction as parasites. Equation (1) gives the mathematical average filtering operation.

$$g_{ij} = \sum_{w=-m}^{m} \sum_{l=-m}^{m} k_{wl} f_{i+w,j+l} \tag{1}$$

where g_{ij} the average filtered image. k is an average filter mask of size 3 by 3. $w = -m \leq 0 \leq m, l = -m \leq 0 < m$, f_{ij} is original image. The parasites appear dark colored objects (purplish) in the microscopic image; however, the visual extent of staining varies due to lighting and imaging conditions. Based on color features, the parasites are very well distinguished in hue-saturation-value color space as compared to the gray level image. Hence the original image is converted to HSV space. Figure 3a–d shows the image after average filtering, hue component image, saturation component image, and value component image, respectively.

2.3 Saturation Separation

The RBCs are lightly stained as mature RBCs do not have nuclei. The parasites consist of chromatin dots which appear as dark objects in the sample image. Saturation channel gives us the degree of color diluted by white light. Considering the images based on intensity, the brighter proportion of the parasites is visible on the light background. Extracting the saturation channel in an image describes color similarities to human eye precipitance. Therefore, to detect a dark object within an RBC, selecting saturation component of the original image gives better understanding of parasites. Histogram of saturation channel helps in finding the threshold value for further processing. Equation (2) denotes the mathematical expression for H component in terms of R, G, B values and Eq. (3) gives a mathematical expression for S component in terms of R, G, and B component, respectively. Figure 4 shows the histogram plot of saturation channel.

$$H = tan \frac{(3(G - B))}{(R - G) + (R - B)} \tag{2}$$

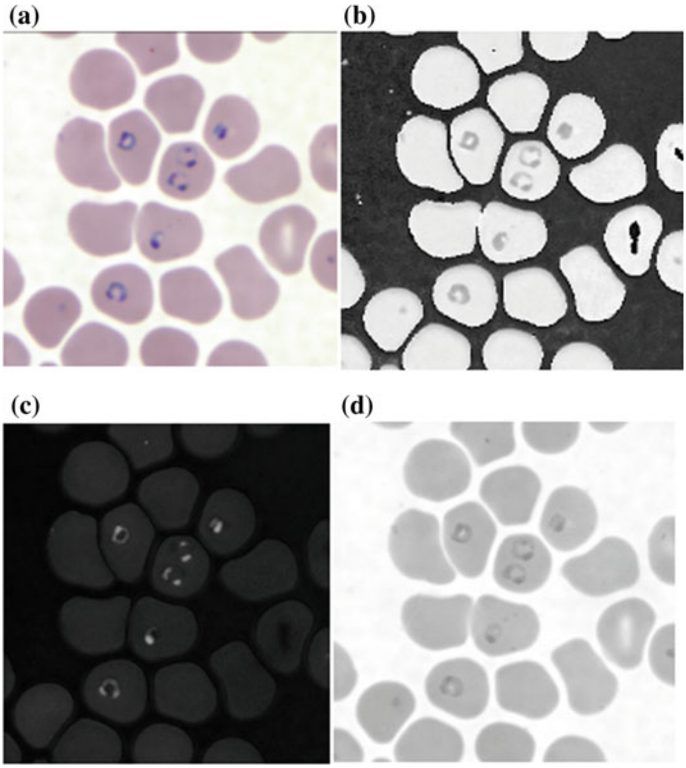


Fig. 3 **a** Filtered image.**b** Hue image. **c** Saturation image. **d** Value image

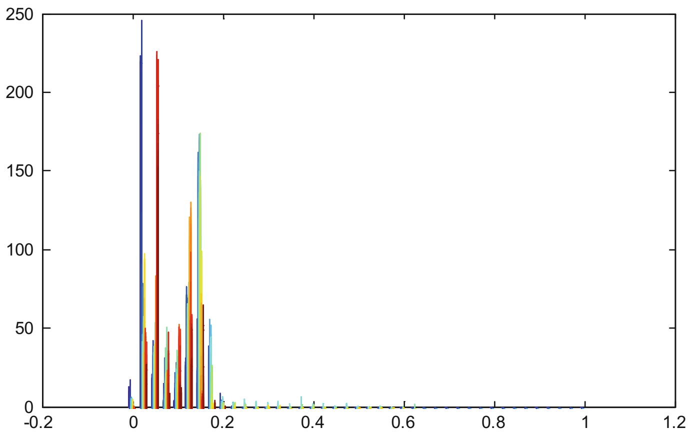


Fig. 4 Histogram plot of saturation channel

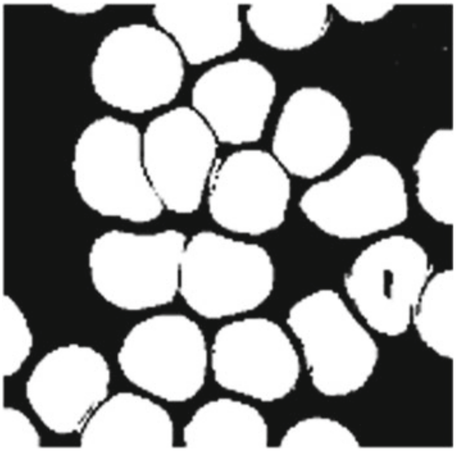


Fig. 5 Triangle's thresholding

$$S = 1 - \frac{\min(R, G, B)}{V} \tag{3}$$

where *R, G, B* is Red, Green, and Blue component images, respectively.

2.4 Triangle's Thresholding

Triangle's thresholding method is performed on the saturation channel. By plotting a histogram of saturation channel, the optimal threshold value is obtained. The triangle's thresholding [11] method constructs a line between the highest and the lowest value of the histogram. Further, the distance is measured between line constructed and the values of the histogram between the minimum and maximum values for all the values denoted by h, where h is the value through which the distance K reaches its highest brightness. The point h at which maximum distance obtained is considered as a threshold. Figure 5 shows the image after triangle's thresholding.

2.5 Analysis of Cells

Applying triangle's thresholding algorithm gives a binary image followed by an opening operation which removes the small unwanted component. The statistical features average, standard deviation of HSV color space and the gray scale are extracted.Support Vector Machine (SVM), a supervised learning algorithm is considered for classification problem of erythrocytes. Binary classification is done to classify whether the random erythrocyte is affected or not. SVM with optimal separating hyperplane plots each point in the n-dimensional space and classifies based on finding the hyperplane which separates the two classes. The mathematical expression for average and standard deviation are mentioned in Eqs. (4) and (5), respectively. The Eq. (6) denotes the expression for linear classification of SVM classifier.

Average:

$$\mu_{H,S,V} = \frac{1}{IJ} \sum_{k=1}^{I} \sum_{l=1}^{J} P_{kl}^{c} \tag{4}$$

Standard deviation:

$$\sigma_{H,S,V} = \left[\frac{1}{IJ} \sum_{k=1}^{I} \sum_{l=1}^{J} \left(P_{kl}^{c} - \mu_c \right)^2 \right]^{\frac{1}{2}} \tag{5}$$

For linear classifier

$$f(x) = w^T x + b \tag{6}$$

where w, b, and x is normal weight, b bias, and support vector, respectively.

3 Experimental Results

The experimentation is conducted on 40 images with different stained images. The samples were obtained from the center for disease control and prevention [12] and Medicare hospital, Ahmednagar. The sample size 256 × 256 is considered as input image for implementation of proposed algorithm. The experimentation performed is implemented in MATLABv2013a environment. Despite the poor slide quality, the system identifies the erythrocytes properly.

$$A = \frac{T_p + T_n}{T_p + F_p + T_n + F_n} \tag{7}$$

$$S = \frac{T_p}{T_p + F_n} \tag{8}$$

$$SP = \frac{T_n}{T_n + F_p} \tag{9}$$

where A, S, SP is overall accuracy, sensitivity, and specificity, respectively. T_p is true positive the classifier identifies an affected cell as an affected cell, T_n is true negative identifies normal as a normal cell; F_p is false positive the affected sample is not identified, F_n is false negative false normal identified. Each and every cell in the original image is checked whether the parasite is present or not. Table 2. shows

Table 2 Comparsion of manual and automatic technique

Patient list	Manual detection of affected cells	Detection by proposed approach
Patient 1	3	3
Patient 2	11	10
Patient 3	5	5
Patient 4	18	18
Patient 5	2	2
Patient 6	2	2
Patient 7	1	1
Patient 8	1	1
Patient 9	2	2
Patient 10	1	1
Patient 11	1	1
Patient 12	1	1
Patient 13	1	1
Patient 14	2	2
Patient 15	4	3
Patient 16	2	2
Patient 17	1	1
Patient 18	1	1
Patient 19	2	2
Patient 20	3	2

Table 3 Comparison of proposed method with existing methods

Method	Specificity	Sensitivity	Accuracy
Diaz et al. [6]	99.7	94	–
Purwar et al. [5]	50–88	100	–
Proposed approach	80	100	85

comparison of results obtained by manual technique and proposed algorithm for identification of infected samples.

Table 3 shows a comparison of the existing methods with the proposed method in terms of specificity, sensitivity, and accuracy.

4 Conclusion

This paper describes the segmentation of RBCs using triangle's method which achieved a good accuracy in identifying an infected RBC and a normal RBC. The proposed work provides a fully automated technique which would help the lab expertise in counting number of parasitic cells, identification of affected cells and will improve management of this febrile illness to ensure the antimalarial medication. Early detection of malaria will be life-saving. Further, we can extend the work by identifying the species and the stages of the affected samples which would definitely help in improving the diagnosis method.

Acknowledgements We are heartily grateful to the Medicare hospital, Ahmednagar for providing us the blood slides images which lead to the implementation of our project and some images were taken from the CDC website.

References

1. World Malaria Report: World Health Organization (2017)
2. Wassmer E (2017) Grau: Severe Malaria: What's new on the pathogenesis front? Elsevier Int J Parasitol 47:145–152
3. Azikiwe, Ifezulike, Siminialayi, Amazu, Enye, Nwakwunite (2012) A comparative laboratory diagnosis of malaria: microscopy versus rapid diagnostic test kits. Elsevier Asian Pac J Trop Biomed 4:307–310(2012)
4. Nugroho HA, son of Ali Akbar, Elsa Herdiana Murhandar E (2015) Feature extraction and classification for detection of malaria parasites in a thin blood smear. In: Proceeding of IEEE conference on information technology, computer, and electrical engineering. pp 197–201
5. Yashasvi, Shah, Clarke, Almugairi A, Muehlenbachs A (2011) Automated and unsupervised detection of malarial parasites in microscopic images. Malar J 1–10
6. Gloria, Gonzalez, Romero (2009) A semi-automatic method for quantification and classification of erythrocytes infected with malaria parasites in microscopic images. Elsevier J Biomed Inf 42:296–307
7. Boray, Dempster, Kale (2009) Computer vision for microscopy diagnosis of malaria. Malar J 9:1–12
8. Minh-Tam Le, Bretschneider, Kuss, Preiser (2008) A novel semi-automatic image processing approach to determine Plasmodium falciparum Parasitemia in Giemsa-Stained thin blood smears. In: Research article BMC cell biology, pp 1–12
9. Kareem S, Morling RCS, Kale I (2011) A novel method to count the RBCs in thin blood films. In: Proceedings of IEEE conference on circuits and systems, pp 1021–1024
10. Kareem S, Morling RCS, Kale I (2012) Automated malaria parasites detection in thin blood films—a hybrid illumination and color constancy insensitive, morphological approach. In: Proceedings of IEEE conference on circuits and systems, pp 240–243
11. Sadeghian, Seman Z, Ramli AR, Kahar BHA, M-Iqbal Saripan (2009) A framework for white blood cell segmentation in microscopic blood images using digital image processing.11:196–206
12. https://www.cdc.gov (centre for disease control and prevention)

Detection and Recognition of Vehicle Using Principal Component Analysis

Kolandapalayam Shanmugam Selvanayaki, Rm. Somasundaram and J. Shyamala Devi

Abstract The idea of detection of moving objects and the concept of classification of moving objects is considered to be the important part of research in video processing and in real-time applications for surveillance and tracking of vehicles. In scientific terms, image processing is said to be any form of signal processing, where the input is an image and the output of image processing may taken be either an image or a set of characteristics or parameters related to the image. The proposed work of the paper is to detect and classify vehicles in a given video. It consists of two modules, first one uses GLOH algorithm for feature extraction and feature reduction. The second module classifies the vehicle from an input video frame using PCA. The final result is obtained by integrating the above-said modules.

1 Introduction

Many of the existing image-processing techniques treat the image as a two-dimensional signal and applies the standard signal-processing techniques to the image. The detection of an object is done by either one method, where the object gets detected in every frame or the object is detected when it appears first in the video [1]. The most predominantly used approach in object detection is using the information exists in a single frame. A common approach for object detection is to

K. S. Selvanayaki (✉)
Department of Computer Science, Concordia University, Chicago,
River Forest 60305, Illinois, USA
e-mail: selvanayaki.ks@cuchicago.edu

Rm. Somasundaram
Department of Computer Science & Engineering, SNS College of Engineering, Coimbatore,
Tamil Nadu, India
e-mail: ilakiasomu@gmail.com

J. Shyamala Devi
Department of Computer Science, SRM University, Chennai, Tamil Nadu, India
e-mail: shyamaladevi@gmail.com

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_97

use information in a single frame. Several features like the shape, color, logo, etc., could be used to detect and classify the vehicle in the video.

In computer vision tasks, GLOH (Gradient Location and Orientation Histogram) is a robust image descriptor which is mostly used. Principal Components Analysis (PCA) reduces the higher dimensionality identified by the descriptor. Therefore, GLOH is a combination of PCA and the SIFT algorithms. GLOH algorithm is performed in two steps initially the keypoints and features are extracted using SIFT algorithm and further dimensionality reduction is achieved using the PCA algorithm. Scale Invariant Feature Transform (SIFT) features are features which are extracted from images that helps to match the different views of the same object. The extracted features are identified to be invariant in considering to scale and orientation and are mostly distinctive of the image. SIFT is not just scale invariant but also rotation, illumination, viewpoint invariant.

2 Principal Component Analysis (PCA)

Principal components analysis (PCA) is one of the classical methods. It provides a sequence of best linear approximations to a given high-dimensional observation; with all the existing techniques, PCA has to be taken as one of the most popular techniques for reduction in dimensionality.

2.1 Principal Component Analysis for Classification

Classification is the process of identifying the set of categories, a new observation belongs by considering the training set of data containing observations for known category membership. The algorithm gives the classification in a concrete implementation is called as a classifier. Sometimes, "classifier" refers to mathematical functions given by classification algorithms, which is mapping input data for this category. Some of the commonly used classification techniques are State Vector Machines (SVM), Neural networks, and Principal Component Analysis (PCA). Some of these techniques are discussed below.

2.2 Existing Approaches for Classification

SVM: Support Vector Machine (SVM) is a discriminative classifier formally defined by a separating hyperplane. It is a supervised learning model which analyze data and recognize patterns that can be used for regression analysis.

Neural Networks: A Neural network consists of a closely connected group of artificial neurons, and process the information in a connectionist approach for computation.

Principal Component Analysis: Principal component analysis is an algorithm which achieves dimensionality reduction. In our proposed system, PCA is used for classification on cars in a frame of video based on its brand. PCA classification aims to maximize between-class separation and aims at accurate classification.

3 Overall System Design

The main focus of this paper is to detect and classify vehicles (cars) in a given video. Gradient Location Orientation Histogram (GLOH) algorithm and [SIFT-PCA] are used for feature extraction and feature reduction. PCA is used for further classification of the vehicle from the input video frame. In this system, there are basically three modules. The SIFT module which consists of key point extraction using SIFT and feature extraction. The next module consists of PCA reduction. The final module is the classification of the vehicle using the PCA algorithm. All the three modules are then integrated to provide the final result and the overall system design is given in Fig. 1.

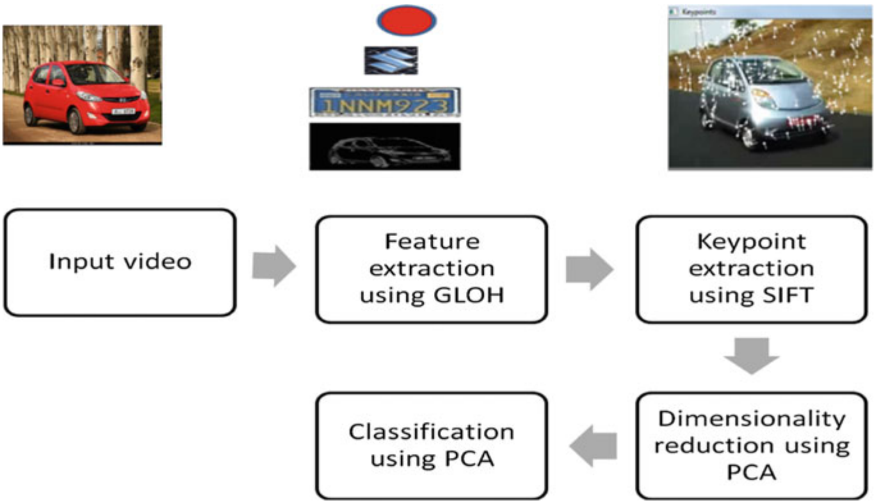


Fig. 1 Overall system design

4 Number Plate Detection and Character Recognition

The aim is to extract the vehicle registration numbers from car images and recognize the same without the use of human intervention for ease of authentication system. The first step is to exactly locate and extract the license plate from car video and the next is to extract the text for further authentication purpose.

Number plate detection is done in two steps.

- Extracting number plate from a video using BLOBs library
- Extracting the characters using Aostsoft Image to Text OCR Converter.

BLOBs library is used for doing computer vision by finding BLOBs on an image, for areas whose brightness is above or below a particular value. It also provides functions and manipulates filtering and extracting results from the extracted BLOBs. In order to extract the numbers in the license plate, Aostsoft Image to Text OCR Converter is used.

4.1 Overall Architecture

The overall architecture for Number Plate Detection is given in Fig. 2.

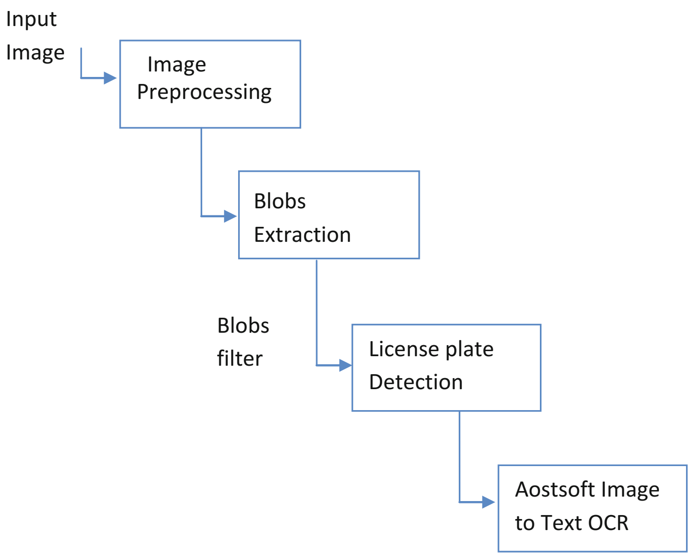


Fig. 2 Overall architecture for Number Plate Detection

Fig. 3 Sequence of preprocessing techniques

4.2 Extracting Number Plate

Number plate extraction involves two steps, namely, preprocessing techniques to reduce noise, BLOBs extraction, and text extraction.

4.3 Preprocessing Techniques

The most important part of image processing is just not detection but the preprocessing techniques that is done before the operation which helps in removing noise and thus reduce errors. Smoothing, morphological operation, and threshold are the three techniques chosen. Input image is first smoothened to remove noise.

Next technique used is TOP HAT, a morphological operation. TOP HAT exaggerates those portions which are lighter than the surroundings. In our case, number plate is lighter region. Next is to apply threshold which makes lighter regions more prominent for BLOBs extraction. OTSU threshold technique is used, which is based on moving the average value of pixels within the image (Fig. 3).

4.4 BLOBs Extraction

BLOB detection is focussed at the detection of points and/or regions in the image which differs in properties like brightness or color compared to their surroundings. Using BLOBs library such regions are extracted. This includes a large collection of BLOBs area of various sizes. The library gives the two basic functionalities namely (1) Extracts eight-connected components also referred as BLOBs in binary or grayscale image and (2) Filters the BLOBs which are obtained to get the interest objects in the image, which used Filter method from CBLOBResult (Fig. 4).

As a result, insignificant BLOB that is either too large or too small is also detected. These BLOBs are filtered based on area and ratio constraints. Ratio is calculated as width/height. It is observed that ratio calculated for the number plate of a car falls in

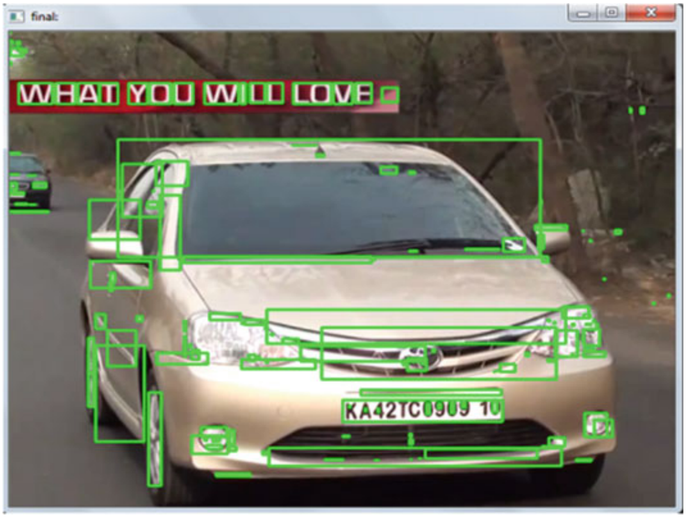


Fig. 4 Result of BLOB detection

Fig. 5 Result of BLOB detection after filtering

the range of 1.8–8 and area less than 8000. Those BLOBs which do not satisfy these conditions are eliminated. Thus, the number plate part is detected (Fig. 5).

4.5 Text Extraction

Now only the number plate is detected. Further, the text is to be extracted and checked whether it is a government car or not. Aostsoft Image to Text OCR Converter is used to extract the text in the number plate. The localized region is given as an input, from which individual characters are extracted which is then saved into a text file. The easy and flexible Image OCR converter is Aostsoft Image to Text OCR Converter. It is much helpful in extracting text from various types of images such as JPG, JPEG, TIF, TIFF, BMP, GIF, PNG, PSD, EMF, WMF, DCX, PCX, JP2, J2K, etc. The authentication is based on the occurrence of the letter "G" or "g". In case if the letter G/g is present then it is classified as Government car else not a government car.

4.6 Color Detection

The one of the important feature is color, which is chosen to uniquely label a group of cars. This system combines two methods to increase the efficiency of color detection over a range of conditions.

Haar cascade classifier to localize the car in the frame or image: A Haar Cascade Classifier which is trained to localize the car region alone in the frame is used to find the car position in the frame. This car position is given as the input for the color detection algorithm which estimates the color of the car.

Find color using the center point of localized region: This module has considered middle pixel and six neighboring pixels and estimates the color of the car using the RGB model. For a video, the color values of the points are averaged to get the exact color [2].

Condition where algorithm fails	Condition where algorithm works properly
(1) When the pixel color represented by the midpoint of the Region of Interest (ROI) is not the proper color of the car	(1) When the pixel color represented by the midpoint of the Region of Interest (ROI) is the proper color of the car
(2) When there is too much of light illuminating the front view, which may give a white dominant shade and not the original color	(2) The system's reliability is not marred by the background like in other methods

4.7 Shape Extraction

This system combines three different methods to extract the shape of the object, namely, BLOBs detection, Finding Contours, and Contour Comparison. Once the

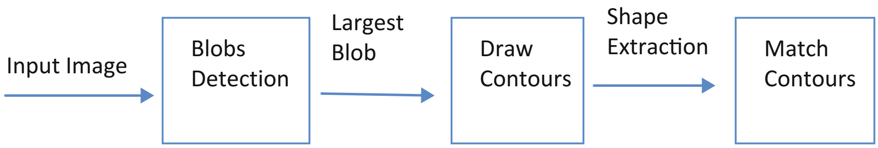


Fig. 6 Flow of shape extraction technique

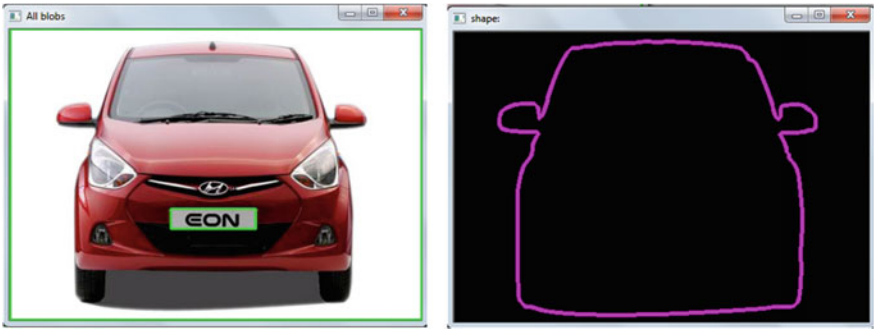


Fig. 7 Result of BLOB detection

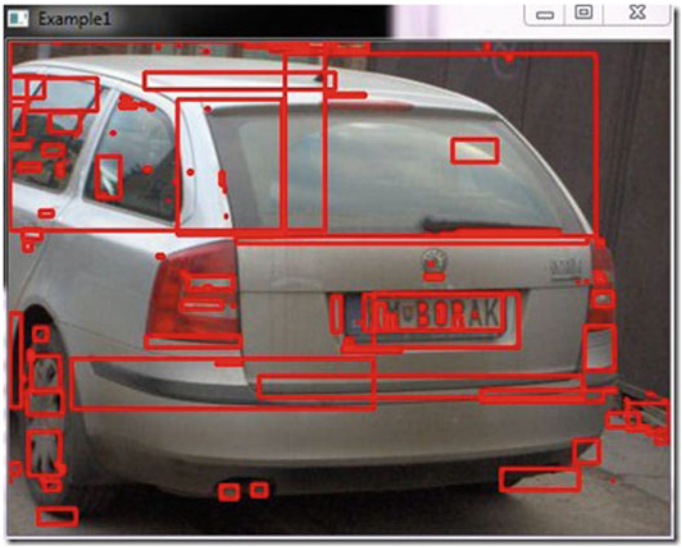


Fig. 8 BLOB detection

shape is extracted, they are compared with templates in order to find the match percentage value (Figs. 6, 7, 8, 9, 10 and 11).

The above algorithms were used to extract user-defined features from the given car in the video. At the end of this phase, a feature vector was formed which described

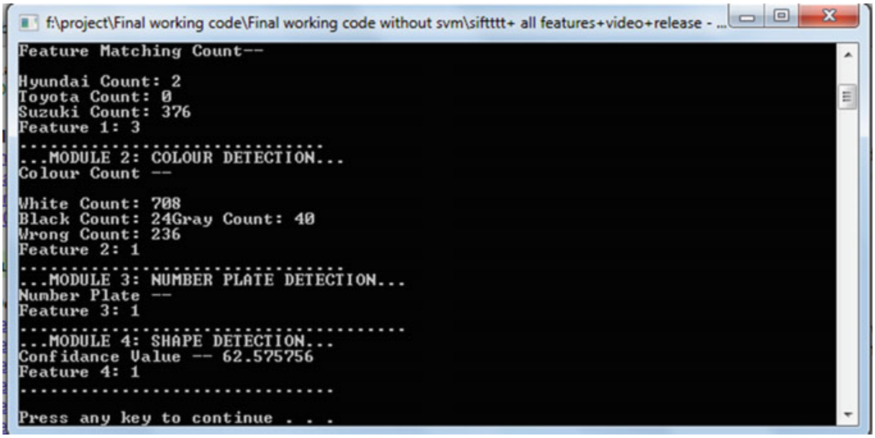


Fig. 9 Output of feature extraction

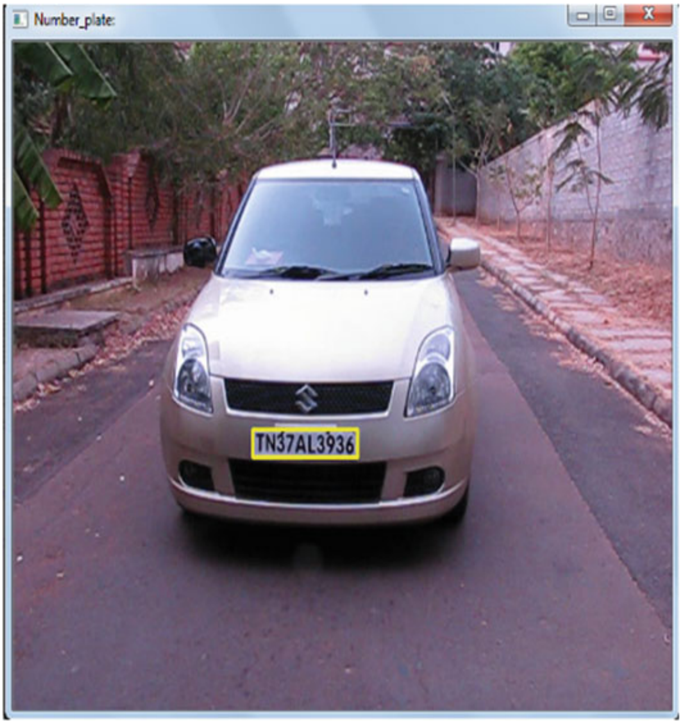


Fig. 10 Output of number plate detection

the features of the given car in the frame. The feature vector has three feature values among which the first value is for the logo the next feature value is for the color and

Fig. 11 Output of shape extraction

the last feature value is for the shape. It was earlier decided that SVM will be used for classification. But since it was suggested that classification could be done using PCA this module was completed at this stage and classification was done using PCA on the reduced frame from the previous phase.

4.8 Steps in PCA Classification

Loading images for training

The identified steps in PCA Classification is as follows:

- Finding the PCA subspace
- Thesising the training images
- Saving the learned face model
- The recognition phase
- Finding the nearest neighbor (Figs. 12 and 13).

5 Conclusion

The classification is done during this phase and classification is done based on the nearest neighbor by finding out the minimum distance. This stage has a training and testing phase where the classification model is learned during the training phase and actual classification is done in the testing phase. This model can be improved by

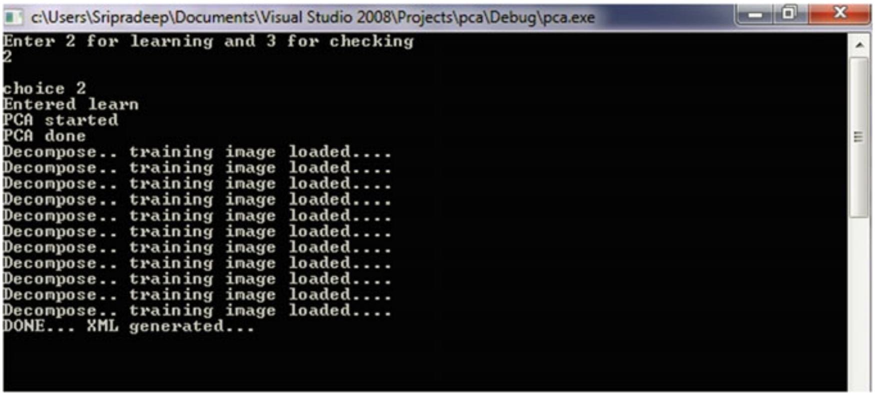


Fig. 12 Training phase output

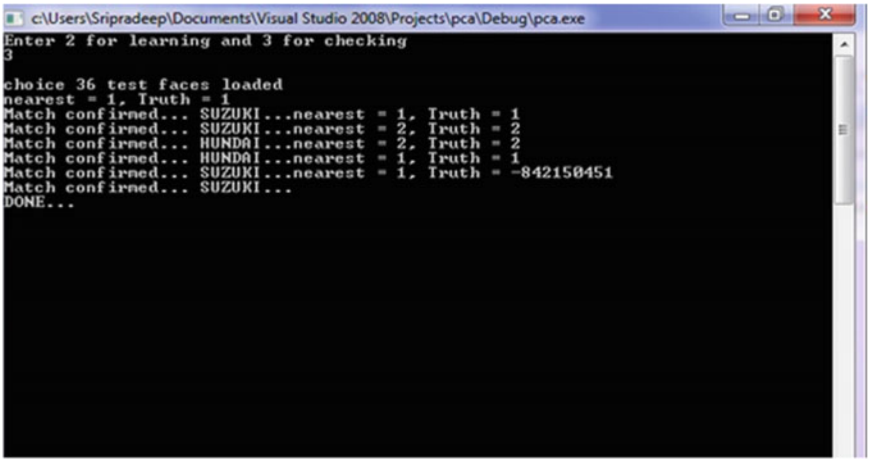


Fig. 13 Test recognition phase

taking other different distance measuring approached to find out the distance. This classification is faster than the already existing method because it is a pixel-based method whereas most of the other methods are feature-based.

There are a few more conditions under which the algorithm misclassifies or fails, which are given as follows

Condition where the algorithm works	Condition where the algorithm fails
(1) The factors like width and height of each input image (parameter 2 of cvCalcEigenObjects(), input image) need to be identical. Also, these images should be single channel	(1) The factors like width and height of each input image (parameter 2 of cvCalcEigenObjects(), input image) need not to be identical. Also these images are not single channel
(2) The factors like width and height of the average image (parameter 9 of cvCalcEigenObjects(), average image) needs to be the same as the width and height of the input images	(2) The factors like width and height of the average image (parameter 9 of cvCalcEigenObjects(), average image) need not the same as the width and height of the input images
(3) The width and height of each eigenvector image (parameter 3 of cvCalcEigenObjects(), output) must also be the same as the width and height of the input images	(3) The width and height of each eigenvector image (parameter 3 of cvCalcEigenObjects(), output) need not be the same as the width and height of the input images
(4) All the images probably need to be single channel	(4) All the images probably need not be single channel

Table 1 Results of various brands of car after classification

Brand	Positive images used	Negative images used	Results obtained for image
Suzuki	10	Zero	It is been to get accepted for the trained set of images
Suzuki	50	10	Worked well for Suzuki cars
Hyundai	5	10	Car detected as some other brand
Hyundai	10	2	Car detected as Hyundai

For a given query image, let's assume:

There are M relevant images of "correct matches" in the collection. Upon execution of the query, it is possible to retrieve N images.

- In that, only R are relevant.
  So, Recall $= R/M =$ Number of retrieved images that are relevant/Total number of relevant images.
- *Precision* $= R/N =$ Number of retrieved images that are relevant/Total number of retrieved images.

The recall is the answer to the question: How close am I to getting *all* good matches? The precision is the answer to the question: How close am I to getting *only* good matches?

In PCA classification precision and recall are inversely proportional.

The classifier works well for Suzuki and Hyundai cars and classifies those cars correctly when there was more number of positive images. The classifier can be extended further for classification of other brands of cars as well. The following are the results obtained for various brands of cars after classification (Table 1).

It can be seen from the results that principal component analysis gives better results with good classification measures.

References

1. Anagnostopoulos CNE, Anagnostopoulos IE, Psoroulas ID, Loumos V, Kayafas E (2008) License plate recognition from still Images and video sequences: a survey. Intell Trans Syst IEEE Trans 9:377–391. https://doi.org/10.1109/tits.2008.922938
2. Butzke M, Silva AG, Hounsell MS, Pillon MA (2008) Automatic recognition of vehicle attribute-color classification and logo segmentation. Hifen, Urugaiana, pp 32–62
3. Cao X, Wu C, Yan P, Li X (2011) Linear svm classification using boosting hog features for vehicle detection in low-altitude airborne videos. In: Proceedings of the 18th IEEE international conference on image processing
4. Zhang C, Chen X, Chen W (2006) A PCA-based vehicle classification framework. In: 22nd international conference on data engineering workshops
5. Lowe DG (2004) Distinctive image features from scale-invariant keypoints. Int J Comput Vision 60(2):91–110
6. Bagarinao E, Kurita T, Higashikubo M, Inayoshi H (2009) Adapting SVM image classifiers to changes in imaging conditions using incremental SVM: an application to car detection. In: Proceeding of 9th Asian conference on computer vision (ACCV), pp 363–372
7. Liu H (2009) Skew detection for complex document images using robust borderlines in both text and non-text regions. Sci Dir, Patt Recogn Lett
8. Kim KK, Kim KI, Kim JB, Kim HJ (2000) Learning-based approach for license plate recognition. In: Proceeding of IEEE workshop on neural networks for signal processing, vol 2, pp 614–623
9. Mikolajczyk K, Schmid C (2005) A performance evaluation of local descriptors. IEEE Trans Patt Anal Mach Intell 27(10):1615–1630
10. Juan L, Gwun O (2010) A comparison of SIFT, PCA-SIFT and SURF. Int J Image Process (IJIP) 3(4)
11. Sivaraman S, Manubhai Trivedi M (2010) A general active-learning framework for on-road vehicle recognition and tracking. IEEE Trans Intell Transport Syst 11(2):267–276
12. Vapnik V, Golowich S, Smola A (1997) Support vector method for function approximation, regression estimation, and signal processing. Advances in Neural Information Processing Systems, Cambridge, MA, MIT Press, pp 281–287
13. Shao Y, Lunetta RS (2010) Comparison of support vector machine, neural network, and CART algorithms for the land cover classification using limited training data points. Environmental Protection Agency, USA
14. Yasser Arafat S, Saleem M, Afaq Hussain S (2009) Comparative analysis of invariant schemes for logo classification. In: Proceedings of international conference of emerging technologies (ICET), pp 256–261
15. Chen Z, Pears N, Freeman M, Austin J (2009) Road vehicle classification using support vector machines. Intell Comput Intell Syst

Convolution Neural Networks: A Case Study on Brain Tumor Segmentation in Medical Care

Jayanthi Prisilla and V. Murali Krishna Iyyanki

Abstract Image segmentation is dividing of medical imaging into parts and extracting the regions of interest. The study involves the images of brain tumors where the tumor part is segmented from the image and analyzed accurately and efficiently. Convolution Neural Network (CNN) has made a tremendous progress in the field of the Medical and Information Technology. With CNN model, one may not be able to reorganize higher risk patients to get immediate aid they require but also communicate through the network to the clinicians, surgeons, eventually improving the standard of patient care in the medical system.

1 Introduction

The convolutional neural networks are making progress in image recognition, image processing, and image segmentation which are essential for medical imaging. A medical database consists of very huge images consisting of X-rays, MRI, and CT-Scans of various ailments. The manual segmentation is time-consuming process and is tedious, and is prone to human error. Automatic segmentation is cheap, reliable, and is scalable; works with large networks with more accurate and better at classification. This segmentation provides a faster and a well-planned treatment for the large-scale research.

J. Prisilla (✉)
Hyderabad, India
e-mail: prisillaj28@gmail.com

V. M. K. Iyyanki
Defence Research and Development Organization, Hyderabad, India
e-mail: iyyanki@gmail.com

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_98

2 Previous Study

Case-Based Reasoning Clinical Decision Support System (CBR CDSS) developed by Yin et al. [1] for diagnosing the Probable Migraine (PM) and Probable Tension-Type Headache (PTTH) with much better accuracy. An AI technique, case-based reasoning is used to solve very near similarity solutions and is accepted to be effective methods of managing implicit knowledge. The study states it comprises two entities: symptoms and diagnosis. The symptoms are of 74 types for which diagnosis is the solution for each case. Genetic algorithm is a search heuristics that is used for optimization and searching problems.

PLA General Hospital China, an International Headache Center from where the testing data sets were collected consisted of 222 former cases. The proposed CBR CDSS provides the approval grounded on the similarity measurement between fresh cases and former cases. The KNN method was considered to be more effective over other algorithms for its simplicity and higher accuracy. The equation is given as similarity $(A, B) = \sum f\ (A_i, B_i) \times w_{i,}$ where A denotes problem case, in every case B indicates the current case, and i is the distinct features from 1 to n, the quantity of attributes is given by n and the similarity function is f of features i in the case A and B, w is the weight of each distinct feature. The efficacy of the proposed system in the work was evaluated using the recall rate, precision rate, f-score, and accuracy. The recall rate and precision is a quantity of completeness and exactness, respectively.

The high degree of exactness in identifying several primary headache; when symptoms overlapped it was impotent to analyze accurately PM and PTTH. Hence the weighted CBR CDSS method was developed to differentiate the two types of probable primary headaches to improve the diagnosis and treatment of headaches [1].

Computer-aided diagnoses developed by Allen J. Moses et al. (2006) enhances the capability of a person's mind, integrates more information (unseen by physicians) and searches instantly far more information of chronic head pain. The major argument over headache whether it is a symptom of a disease or an illness is still unknown. This led to the development computer software, computer-aided diagnoses, which help the physician to diagnose the headache accurately and give the treatment accordingly.

The results of the data analysis emphasize the need for thorough exploration of computer-aided diagnoses. A computerized database appears to be reasonable and useful for data gathering. A large user database is needed for giving individual diagnoses statistically. Statistical analysis of databases is as large as the Internet is capable of generating relevant information on more headache sufferers. However, currently such evidence-based standards do not exist in statistics, but data mining and information technology offer evidence-based and potential for exploration [2].

Diagnostic prediction model for the vestibular disease developed Grill et al. [3] and was implemented into a mobile device application that may assist PHP with their clinical decisions. The most common complaints are vertigo and dizziness. The signs and symptoms of vertigo and their complex interactions are based on well-performing prediction model. The data is collected from over 3000 patients per year at German Center for Vertigo and Balance Disorders. Patients are put to a series of questions to determine the type of vertigo and dizziness. Accuracy is defined by the ratio of all correct classifications, i.e., true positives (t_p) and true negatives (t_n) over all collected samples. **Overall accuracy** is equated to $\mathbf{(t_p + t_n)/(t_p + f_p + t_n + f_n)}$.

Four steps are involved in machine learning workflow:

1. Cleansing and preparing data for training, checks redundant features and replaces missing data.
2. Selection of model, used to predict the accurate vertigo diagnosis from a number of independent variables based on any one algorithm—multi-class prediction task, namely, multi-class decision forests, an artificial neural network with one hidden layer and a multi-class logistic regression algorithm.
3. Model implementation and training, the modules are manually placed, connected, and configured. Each model provides different configuration parameters.
4. Evaluation, selection, fine tuning of models and publication of each model provides different parameters to optimize the accuracy. The multi-class decision forest is most suitable model for that task based on varied with the parameters of this model, number of decision trees, depth of the trees and the number of random splits of the trees.

The study involves 688 patients for sociodemographic and diagnostic information. With the overall accuracy of the multi-class decision forest was 0.62, of the multi-class neural network 0.33, and of the multi-class logistic regression 0.18. The cloud machine learning platform combines a high usability with breadth of the evaluation results and their visualization [3].

Mirarchi et al. [4] suggested a framework that implements pathology distribution and classification, it involves gathering and investigating data inflowing from vestibular system to attain supplementary information and support physicians in diagnosing. The study of vestibular functions is to recognize and distinguish vestibular disorders that are liable in human pathologies such as head tilt, asymmetrical ataxia, or nystagmus. The test named Vestibular Evoked Myogenic Potentials (VEMPs) is applied with external electrodes to evaluate the vestibular system that can be performed in cervical and ocular zone to measure the myogenic potentials called, respectively, cervical VEMP (cVEMPs) and ocular VEMP (oVEMPs) potentials. The methods such as (1) neural networks, (2) decision trees, (3) genetic algorithms, and (4) nearest neighbor method are adopted for healthcare. The analysis of health data is done by Bayesian-based methods. The statistical dependencies for gene data are carried out by Bayesian networks. The Entity Relational Diagram was used for clinical information such as the database of patient, examination, treatment, diagnosis and device data and information. PostGreSQL, with spatial extension, is used for database instance and further the system was developed by using a web application and graphical user interface. The probability of an experiment was analyzed based on Bayesian methods. The libraries include R software for statistical computing and graphics and using Weka module, the results are compared for auto-learning.

The proposed study enhances to get related and valuable information of vestibular pathology to the sex. The development of vestibular pathology was carried out with the analysis performed, improving the quality and the quantity of data [4].

Maximal frequent item-set Algorithm (MAFIA) is an algorithm used by Nishara Banu and Gomathy [5] for mining highest frequent itemsets from a database. The transactional database is stored efficiently as a series of vertical bitmaps using MAFIA. Decision trees are powerful for classification and prediction using C4.5 algorithm. The method of operation of k-means algorithm helps achieve its name. In data mining, clustering is used to identify the interesting patterns in a specified dataset. In medical imaging, the k-means algorithm is used for unordered method of defining clusters.

The study involves predicting heart diseases and the heart disease database is preprocessed effectively by deleting corresponding records and inserting missing values. The well-ordered preprocessed dataset is collected by K-means algorithm further processed with the K value of 2. The predicted values of heart attack and prescription ID with their levels are calculated. The data is trained to foresee the heart attack level and shows information gain for the efficient heart attack level using the C4.5. High accuracy, high precision, and recall metrics are achieved [5].

Health care is unstructured and huge. Hence, C4.5 decision tree has high values over k-mean-based MAFIA and k-mean MAFIA with ID3. The study shows efficient fragmenting and extracting significant forms from the heart attack database warehouses for the prediction of heart attack.

Halicek [6] has developed an automated classification of normal and cancerous, head and neck tissue using deep convolutional neutral networks (CNNs). The deep learning has the potential to be implemented into a tissue classifier, fully trainable on a database of hyperspectral images from tissue specimens that can produce near real-time tissue labeling for intraoperative cancer detection. A CNN was implemented using tensor flow to classify the spectral patches as either normal or cancer tissue. The neural network architecture consisted of six convolutional layers and three fully connected layers. The patch size used was 10×10, and the kernel size used for convolutions was 3×3. The output of each convolutional layer is $10 \times 10 \times N$, where N is the number of filters in the convolutional layer. The final layer, i.e., softmax, generates a probability of the pixel belonging to either class. The CNN classification performance was evaluated using leave-one-patient-out external-validation to calculate the sensitivity, specificity, and accuracy.

The experimental results show that the CNN has the potential for use in the automatic labeling of cancer and normal tissue using hyperspectral images 81% sensitivity, 78% specificity, and 80% accuracy, which could be useful for intraoperative cancer detection. The proposed technique is fast and does not require any further post-processing to enhance the results [6].

Prisilla et al. [7] discusses about the role of the deep learning has an application of the artificial intelligence helping to predict the early detection of the disease. The outstanding results in the image and speech recognition of the deep learning have proved over other techniques at predicting the potential drug molecules and analyzing particle accelerator data. The manual data analysis coupled with few methods is used for the unpredictable growth of medical databases for efficient computer-assisted analysis. The powerful tools for quality control like artificial neural networks (ANNs) are used in modeling complex relationships in multifaceted processes [7]. The results showed that the deep learning technique can reduce the error in diagnosing the disease, predict the result faster and provide indispensable treatment for the patient's life-saving.

Prisilla [8] identifies that the quality transmission is required to maintain the standardization, reliability, and correctness in the medical network with the packet loss. This medical network is referred to as butterfly network. The concept of butterfly network is to bring all the physicians and engineers to a unique platform where any clinic can access the services of the equipment by sharing among them. Thus, cloud computing helps in improving diagnoses. The butterfly network topology is based on mesh topology, where nodes take a square form. This butterfly network topology includes $\mathbf{(K+1)2^K}$ nodes organized in $K+1$ ranks, each $\mathbf{n = 2^K}$ nodes. Any specific length is not mentioned for the edges, as the nodes keep increases and then the physical distance between them increases. In 3D network, communication time increases as the network grows and more over the edge length increases with number of processors. The source sends the packets to destination by avoiding the interference or packet loss with the help of network coding. Encoding the contents and forwarding the packets to help to achieve maximum bandwidth. The key purpose of networking coding is it delivers efficiently, ease of management, and resilience to network dynamics. Every node is anticipated to be a medical care center linked with each other. The information is passed in the form of packets from one medical care to another using cloud-based services.

In this study, the performance of the network is increased with the reduction of packet loss and the packet loss is identified very easily. The three causes for packet loss are (1) network link failure (2) when the buffer overflow occurs (3) the packet does not arrive at its destination due to congestion.

The butterfly network has high security with few resources and aims in reducing the cost and network diameter [8].

3 Convolution Neural Network

Captivating an image as input and labeling it as output is referred as image recognition. When a machine reads an image, it reads as an array of pixel values, each value between ranges of 0–255 shown in Fig. 1.

Convolution is a process of combining two functions by multiplying them. To convolve means to roll together. Basically, a convolutional network implements a kind of search. Convolutional networks are used for image detection and take many searches over a single image in horizontal lines, diagonal ones, vertically several times as a visual element is required. A kernel is defined as a matrix of numbers which is used in image convolutions. Various sized kernels comprising different patterns of numbers produce different results and the size of a kernel is arbitrary but 3×3 is used maximum.

In Fig. 2, the INPUT image [$32 \times 32 \times 3$] will hold the raw pixel values, an image of width 32, height 32, and with three color channels red, green, and blue. The result

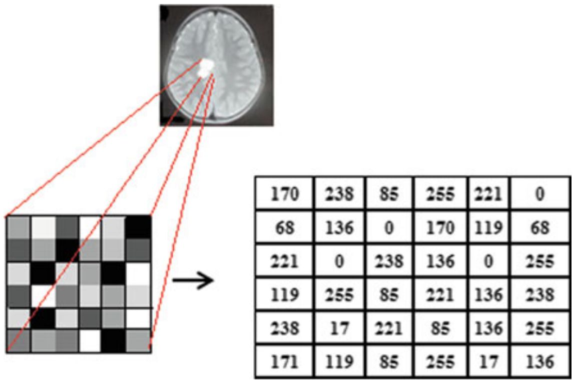


Fig. 1 How machine reads an image in pixels

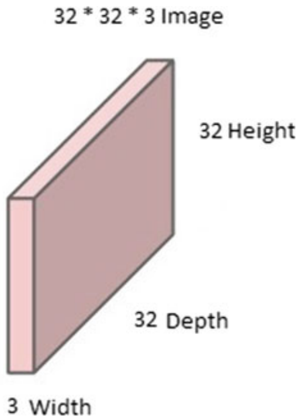


Fig. 2 $32 \times 32 \times 3$ image

of taking a dot product between the filter and small $5 \times 5 \times 3$ chunk of the image ($3 \times 3 \times 3 = 75$ − dimension dot product + bias) as in Fig. 3. When the filter slides around the input image or convolving, it is multiplying the values in the filter with the original pixel values of the image, i.e., element-wise multiplications.

A Conv Net has many layers. The three principal layers of Convolution Neural Network are as follows:

1. CONV Layers
2. ReLU Layer (Rectified Linear Unit) in Fig. 5
3. Subsampling Layers (Fig. 4).

The output of neurons are connected to the inputs and are calculated by CONV layer, each computing a dot product between their weights and a small region connected to the input volume.

ReLU layer applies an element-wise activation function, such as the max(0, x). Thresholding at zero leaves the size unchanged [$32 \times 32 \times 12$]. POOL layer performs a down-sampling operation along the spatial dimensions (width, height), resulting in volume such as [$16 \times 16 \times 12$] and used widely in CNN.

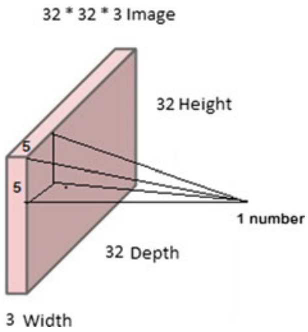


Fig. 3 Applying 5×5 filters

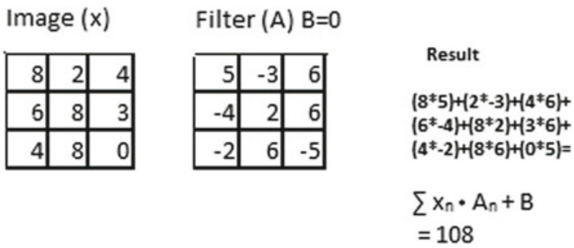


Fig. 4 Filtering

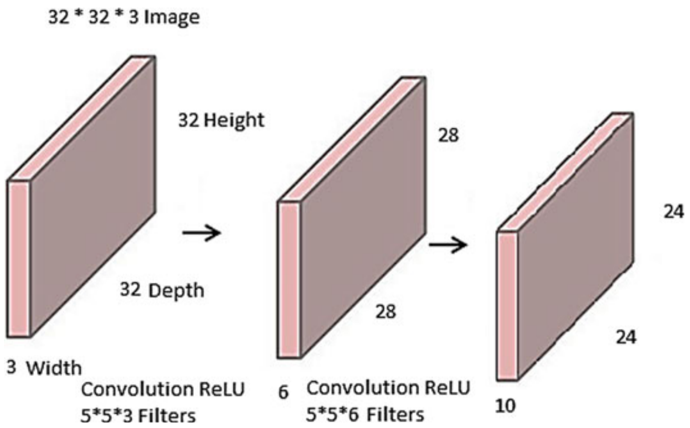


Fig. 5 Conv–ReLU layers in sequence

Max pooling is a technique of subsampling, the subsampling is to get an input representation by reducing its parameters of the network, which helps in reducing overfitting. Max pooling takes the highest pixel value from a region depending on its size. The pooling size is 2×2 pixels as in Fig. 6 [9].

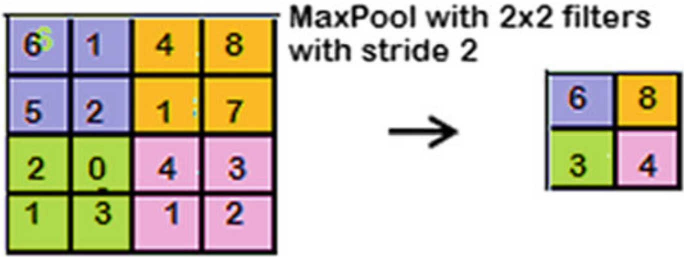


Fig. 6 Maxpooling

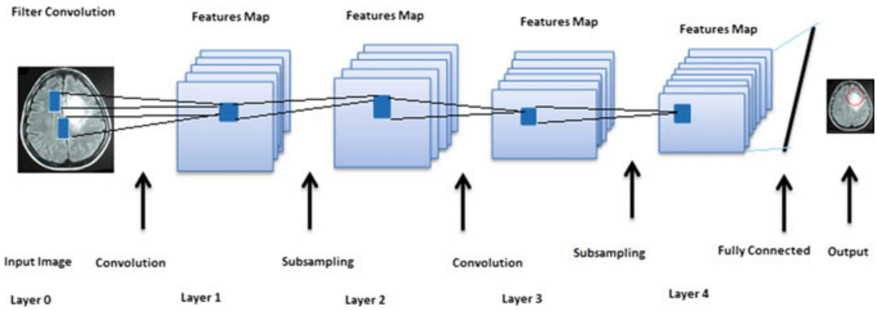


Fig. 7 Fully convolutional neural network

FC, fully connected layer of neurons at the end of CNN will compute the resulting size [$1 \times 1 \times 10$] and it is multi-layer perceptron that uses a softmax activation function in the output layer. Every neuron in the previous layer in the fully connected layers is connected to every neuron on the next layer. The image is segmented to the proper labels after retrieving all of the features from each image as shown in Fig. 7.

4 Results and Discussion

4.1 Image Processing

The elements of image analysis are divided into three areas first, low-level processing, second, intermediate level processing, and last, high-level processing. This research concentrates on the intermediate-level processing that deals with extracting and describing the features in an image resulting from a low-level processing. It involves segmentation and description.

Gray-level slicing, the simple intensity transformation deals with the application which includes enhancing features like flaws in X-rays images, MRI, and CT-scans. In two approaches, one approach deals with the display of high values for all gray levels

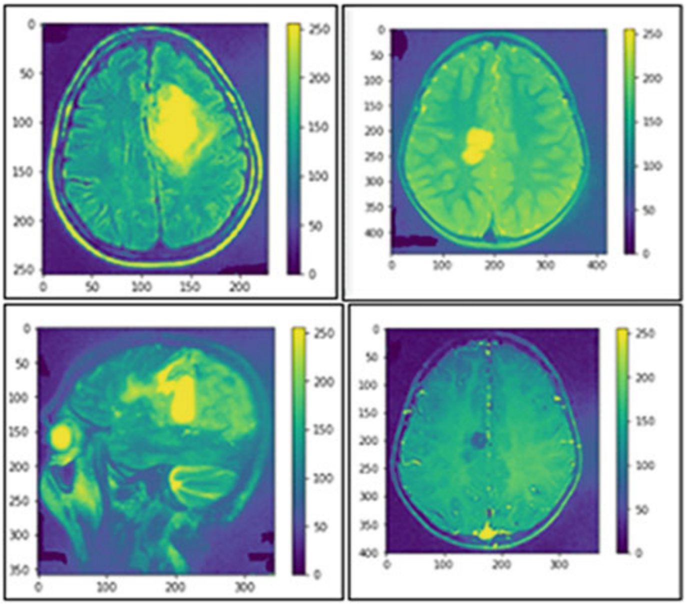


Fig. 8 Tumor part is highlighted

in the range of interest and low value for all the other gray levels. This transformation results in a binary image. Based on the transformation, the second approach brightens the desired range of the gray levels of the image but preserves the background and gray-level tones in the image.

Image subtraction, the difference between two images say $p(x, y)$ and $q(x, y)$ is represented as $R(x, y) = p(x, y) - q(x, y)$, the difference between all pairs of corresponding pixels from p and q. This important feature is used in segmentation and enhancement as displayed in Figs. 8 and 9 [10].

4.2 Image Segmentation

The process of partitioning a digital image to multiple segments, the aim of segmenting an image is to give meaningful information. The segmentation assigns a label to every pixel so that every pixel of same characteristics holds the same characteristics.

The proposed CNN experiment was conducted on the X-ray image of 100 patients having brain tumors which were collected from Gandhi Hospital, Secunderabad, India. The proposed CNN is implemented with the Spyder, a powerful interactive development environment for the Python language. The experiments are performed on a PC with an Intel Core i5-4300 M CPU @ 2.60 GHz processor and 8 GB RAM. On average, the CNN takes about 1 s processing time for each image.

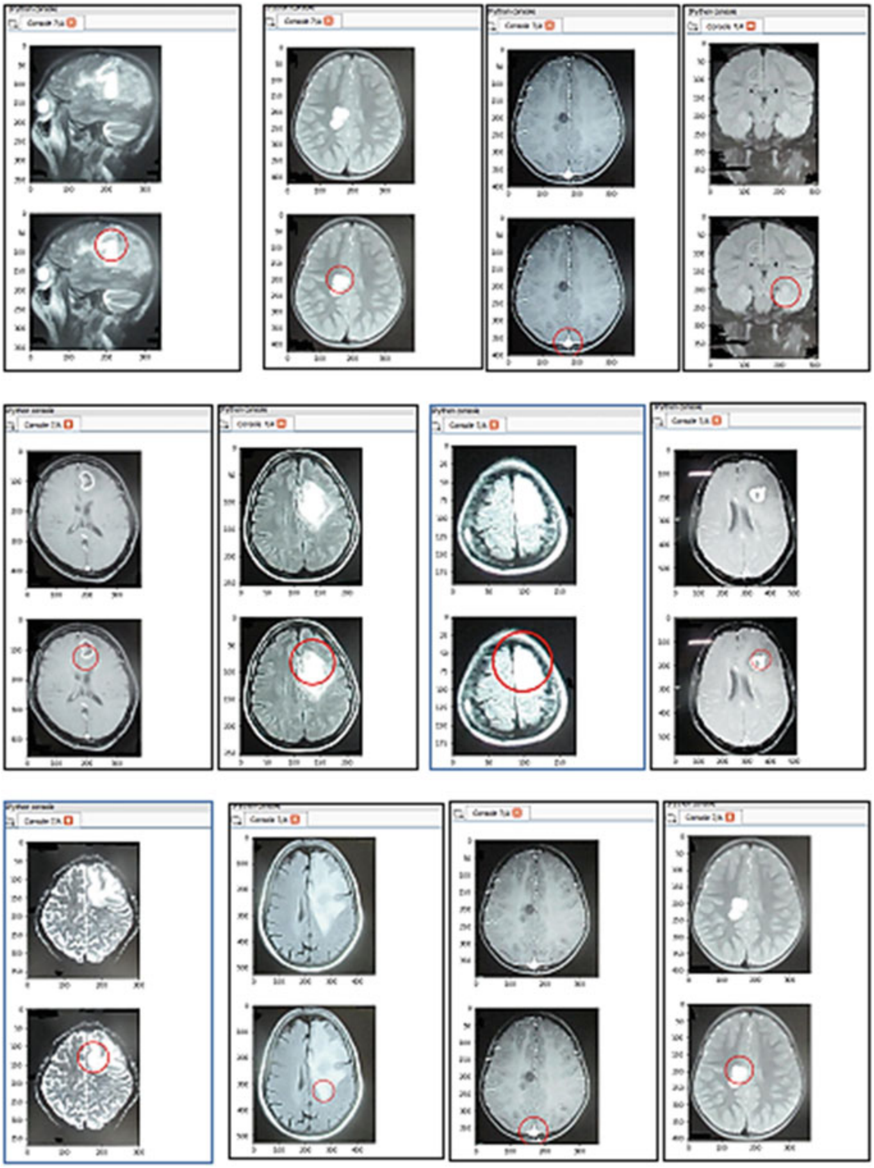


Fig. 9 Outputs of CNN for detecting the tumor part in the brain tumor

5 Conclusions

Fully convolutional networks have a rich essence of models, of which classification convnets are most significant. Extending these classification nets to segmentation, and improving the architecture with multi-resolution layer combinations improves simplifying and speeds up learning. The integration of image segmentation is such that the network can learn certain specific feature than any macroscopic features, and moreover inclusion of more abnormal examples which are very variant and unusual. Hence, the CNN gives more accurate segmentation result which helps in understanding the tumor part in the images of the brain thus helping the surgeon to diagnose it faster and for further analysis. The research is still in progress to recognize the technology growth in the field of medicine.

Conflicts of Interest Prisilla Jayanthi is the principal investigator of the CNN study and involved in script making. Dr. I. V. Murali Krishna is a keen guide in directing the research work and giving the novel ideas. The X-ray reports were collected from the Gandhi Hospital, Secunderabad, India and thanks for their support.

References

1. Yin Z et al (2015) A clinical decision support system for the diagnosis of probable migraine and probable tension-type headache based on case-based reasoning. J Headache Pain 16:29
2. Moses AJ et al (2006) Computer-aided diagnoses of chronic head pain: explanation, study data, implications, and challenges. J Craniomandibular Pract
3. Grill E et al (2016) Developing and implementing diagnostic prediction models for vestibular diseases in primary care. In: Exploring complexity in health: an interdisciplinary systems approach
4. Mirarchi D et al (2016) Applying mining techniques to analyze vestibular data. Elsevier Procedia Comput Sci 98(2016):467–472
5. Nishara Banu MA, Gomathy B (2013) Disease predicting system using data mining techniques. Int J Tech Res Appl. e-ISSN: 2320-8163
6. Halicek M et al (2017) Deep convolutional neural networks for classifying head and neck cancer using hyperspectral imaging. J Biomed Opt 22(6):060503
7. Prisilla J et al (2017) Deep learning in oncology—a case study on brain tumor. Int J Cancer Res Ther. ISSN: 2476-237
8. Prisilla J (2016) The butterfly network in medical care using clouds services. ICTBIG IEEE. https://doi.org/10.1109/ictbig.2016.7892655. Electronic ISBN: 978-1-5090-5515-9
9. Chen K, Seuret M (2017) Convolutional neural networks for page segmentation of historical document images
10. Gonzalez RC, Woods RE, Digital image processing. Pearson Education Asia

Vision-Based Algorithm for Fire Detection in Smart Buildings

Patel Abhilasha Paresh and Latha Parameswaran

Abstract Recent technological advancement has opened the space for a gradual increase in the number of smart buildings. Public safety and security has becomes a matter of concern with such a development, especially in areas of fire accidents. The conventional fire detection system usually worked on sensors and takes time for fire detection. This work presents an early fire detection system that unlike conventional fire detection system is cost-effective with high fire detection rate. The proposed algorithm uses features like color, increase in area and intensity flicker for early detection of fire. Segmentation of fire colored regions is done with the help of $L^{*}a^{*}b^{*}$, YCbCr, and RGB color space. Analysis of fire, i.e., fire area, its spread, temporal information, direction of the fire, and its average growth rate are measured using optical flow and blob analysis. Accuracy and F measure are used to evaluate the accuracy of the proposed system. Experimental results show that the average accuracy of the system is above 80% which is more promising in a video.

1 Introduction

The social progress, economic development, and technological advancements have increased the construction of buildings across a very large space. With such advancements, public safety and security have become a matter of concern especially in the case of fire accidents. Fire could cause huge economic and ecological damage and most importantly endangering people's life. Thus, such loss due to fire can be avoided or reduced with the help of an efficient and early fire detection system. Early detection of fire would limit the loss to minimum. Conventional fire detectors work on the basis of certain fire features like smoke, heat, or radiation. These fire detectors are

P. A. Paresh · L. Parameswaran (✉)
Department of Computer Science and Engineering, Amrita School of Engineering,
Amrita Vishwa Vidyapeetham, Coimbatore, India
e-mail: p_latha@cb.amrita.edu

P. A. Paresh
e-mail: cb.en.p2cvi16005@cb.students.amrita.edu

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_99

usually based on infrared sensors, optical sensors, and ion sensors that respond to fire features. The reliability and the positional distribution of these sensors play an important role in early and efficient fire detection. Thus, early detection of fire in such a scenario seems practically impossible. The conventional system requires dense distribution of sensors, large detection time, costly, and thus fails to meet the needs for fire detection in large areas. Vision-based fire detection system not only overcomes these concerns but it also unlike conventional fire detection system facilitates earlier detection of fire and provides information regarding the location, direction, spread, and growth rate of fire.

This paper is structured as follows: Sect. 2 investigates the survey on the works in visual fire detection domain. The proposed work is described in Sect. 3. It discusses the architecture diagram and algorithms used in this paper. Architecture diagram discusses the workflow of the proposed work while algorithm section explains various techniques used in this paper. In Sect. 4, results and analysis of the proposed work based on different data set are discussed before drawing conclusion in Sect. 5.

2 Related Works

With the increasing concern of safety and security of life and property, there is a considerable increase in the development of visual fire detection systems that serve the need for fire detection in smart buildings. These systems detect fire with the help of color, shape, motion, geometry, and texture characteristics of fire. Color detection and dynamic feature analysis form the basis for most of the vision-based fire detection systems. A survey of key techniques is discussed here.

Celik and Demirel [1] proposed a color model that checks for the fire pixels using YCbCr color model which distinguishes luminance from chrominance than any other color model like RGB and $L^*a^*b^*$. Zhang et al. [2] considering the flicker characteristic of a fire, proposed a forest fire detection system based on contours of fire flame. This work combines two methods, i.e. wavelet which is basically used to detect fire, pixels is combined with FFT that describes the contour of the fire.

$L^*a^*b^*$ color model is used in [3] for fast and efficient fire detection system. Two improved features like color and motion are used in [4] to propose an improved probabilistic approach. In [5] color information, area size, coarseness of the surface, roughness of the boundary, and so on, are taken into consideration by Borges et al. to evaluate the behavioral change by observing the changes in between the frame with the help of low-level feature. Bayes classifier is used to detect fire. Considering the flicker property of a fire, Toreyin et al. [6] uses wavelet transform for detection of flickering pixels that indicate the presence of flames along with the Hidden Markov Models. Bohush and Brouka [7] use contrast and motion features to detect smoke while color and texture features are used for detection of flame in video.

3 Proposed Work

This work proposes an early fire detection system using color, flicker, area, and motion characteristic of fire for various online videos. Color segmentation is used to extract all the regions resembling the fire color and intensity. The color models used are RGB, YCbCr, and $L^*a^*b^*$. Morphological operations are used as post-processing to fill in for the fire pixels not detected due to the use of different color models. Flicker image is obtained by segmenting the high flicker feature regions. Optical flow is used to measure the speed and the direction of fire. Blob analysis, on the other hand, describes the height, width and the area of the detected fire.

3.1 Architecture Diagram

The architecture of the proposed algorithm is shown in Fig. 1. Two consecutive frames from input video are given to color segmentation module that extracts all colors ranging from yellow to red [1, 3, 8–10]. The empty spaces in extracted region are then filled using dilation operation. The rate of change in the area between the two consecutive frames is then measured. The frame is considered for fire detection if change in the area is greater than the threshold value. On the other hand, flicker image is measured by considering all the frames in 1s [11]. Both the images are then masked to get a fire image. The analysis of this fire is done with the help of Luckas Kanade optical flow and blob analysis [5, 12, 13].

3.2 Algorithm

Segmentation Using Flicker Intensity flickering feature is one of the important characteristics in fire detection. Since, positions of pixel around fire regions may or may not be a flame pixel in short time interval, thus having strong intensity differences. A pixel is considered to be fire if conditions in Eqs. (1) and (2) are satisfied [11].

$$D_{x,y} = \sum_{t=1}^{100} I_{x,y}(t) * \mathrm{diff}_{x,y}(t) \quad (1)$$

$$D_{x,y} \geq \text{threshold} \quad (2)$$

High flicker increases the value of $D_{x,y}$ as in Eq. (1). The distinguishing factor that separates fire from non-fire is the second condition as in Eq. (2). A pixel is said to be fire if $D_{x,y}$ is greater than threshold otherwise it is just a disturbance pixel.

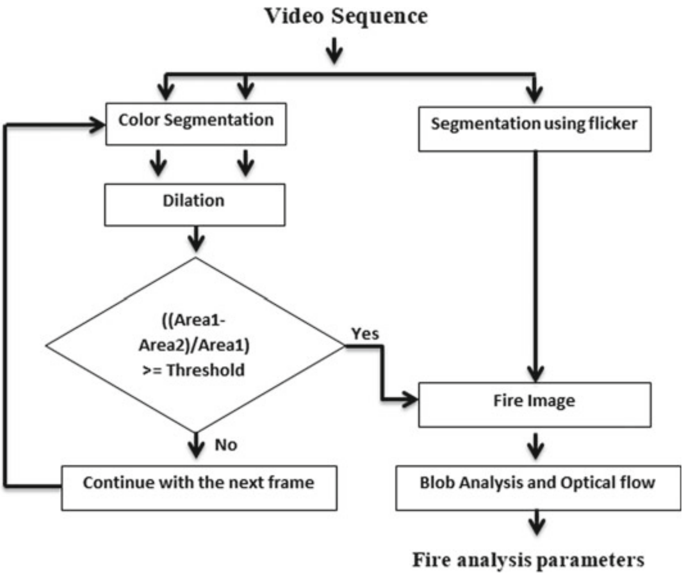


Fig. 1 Architecture of the proposed algorithm

Segmentation Using Color Fire can be characterized by color. Fire being high in illumination and color in the range of yellow to red, we can use this to characterize the fire pixels and detect its existence. Rule-based color segmentation for fire color regions is done using three color models, i.e. $L^*a^*b^*$, YCbCr, and RGB [14]. $L^*a^*b^*$ and YCbCr are chosen because of its ability to extract the illumination and chrominance separately for the segmentation of high-intensity regions. YCbCr is used to detect very high-intensity yellow color regions. $L^*a^*b^*$ color space, on the other hand, is used to detect orange and darker orange regions. RGB is used because most of the visible video cameras capture videos is in RGB format which makes our segmentation computational less complex. RGB is used to extract red regions of fire.

Lab Color Model Color of the fire usually ranges from yellow to red. $L^*a^*b^*$ color model is basically used to detect orange and darker orange fire. For an image to check for the existence of fire in it using a $L^*a^*b^*$ color model, its statistical measures like average can be calculated using the Eqs. (3), (4) and (5) [3, 9, 10].

$$L_{\mathrm{m}} = \frac{1}{N}\sum_{x}\sum_{y} L^*(x, y) \tag{3}$$

$$a_{\mathrm{m}} = \frac{1}{N}\sum_{x}\sum_{y} a^*(x, y) \tag{4}$$

$$b_{\mathrm{m}} = \frac{1}{N}\sum_{x}\sum_{y} b^*(x, y) \tag{5}$$

where

L^*_m, a^*_m and b^*_m	are average values of the L^* (luminance),
a^*, b^*	(chrominance), respectively,
N	is the total pixels in a frame and (x, y) is spatial pixel location in an image.

$L^*a^*b^*$ color model assumes that the regions of fire in a frame is the most luminous region and also is near to the color red. Thus, considering these conditions, fire rules defined in Eqs. (6), (7), (8) and (9) are formulated [3]

$$R1(x, y) = \begin{cases} 1, & L^*(x, y) \geq L^*_m \\ 0, & \text{Otherwise} \end{cases} \tag{6}$$

$$R2(x, y) = \begin{cases} 1, & a^*(x, y) \geq a^*_m \\ 0, & \text{Otherwise} \end{cases} \tag{7}$$

$$R3(x, y) = \begin{cases} 1, & b^*(x, y) \geq b^*_m \\ 0, & \text{Otherwise} \end{cases} \tag{8}$$

$$R4(x, y) = \begin{cases} 1, & b^*(x, y) \geq a^*_m \\ 0, & \text{Otherwise} \end{cases} \tag{9}$$

A pixel is said to be a fire pixel if it satisfies all the conditions by Eqs. (6), (7), (8) and (9).

RGB Color Model RGB color model extracts red color information of the image. Considering the RGB value of an image, a pixel is defined as a fire if its R channel intensity value is more than G channel value, G channel intensity value is more than B channel value. Since red color is more prominent than green and blue in fire, we can also say that for a pixel to be fire it needs to have more red intensity than its average red intensity. So mathematically, for a pixel to be a fire, it must satisfy Eqs. (10), (11) and (12) [8].

$$R_m = \frac{1}{N} \sum_x \sum_y R(x, y) \tag{10}$$

where

R_m is the mean of the R channel.

$$R1(x, y) = \begin{cases} 1, & R(x, y) > G(x, y) > B(x, y) \\ 0, & \text{Otherwise} \end{cases} \tag{11}$$

$$R2(x, y) = \begin{cases} 1, & R(x, y) \geq R_m \\ 0, & \text{Otherwise} \end{cases} \tag{12}$$

YCbCr Color Model YCbCr is an efficient color model that responds to high luminous yellow and brighter yellow regions. In YCbCr color model, for fire pixel, following relation between Y, Cb, and Cr channels holds true Y(x, y) ≥ Cr(x, y) ≥ Cb(x, y) [1]. A pixel is said to be a fire if it satisfies Eqs. (13), (14), (15), and (16).

$$R1(x, y) = \begin{cases} 1, \ (R(x, y) \geq G(x, y)), (G(x, y) > B(x, y)) \\ 0, \ \text{Otherwise} \end{cases} \tag{13}$$

$$R2(x, y) = \begin{cases} 1, \ (R(x, y) > (190)), (G(x, y) > 100), (B(x, y) < 140) \\ 0, \ \text{Otherwise} \end{cases} \tag{14}$$

$$R3(x, y) = \begin{cases} 1, \ Y(x, y) \geq \text{Cb}(x, y) \\ 0, \ \text{Otherwise} \end{cases} \tag{15}$$

$$R4(x, y) = \begin{cases} 1, \ \text{Cr}(x, y) \geq \text{Cb}(x, y) \\ 0, \ \text{Otherwise} \end{cases} \tag{16}$$

Blob Analysis Blob is a patch of connected pixel having similar properties to form a single structure. In case of binary image, background is denoted as zero whereas every non zero value represents the object. The main aim of blob analysis is not to detect fire rather analyze the detected fire, i.e. area covered by fire, its height and width [13]. A bounding box is drawn over the detected fire regions. Time at which fire is detected can also be obtained by considering the frame rate and frame for which fire was detected.

Optical Flow Optical flow is used in this work to track the features and measure its displacement between the two consecutive frames [5]. It measures the average rate and the direction of fire. The optical flow algorithm (Lucas–Kanade) [12] is based on three assumptions as follows:

Brightness Constancy It assumes that the brightness of each pixel remains alike from frame to frame. Considering the grayscale images, the change in the intensity of the pixels is always lesser than the threshold such that pixels are tracked properly frame by frame.

$$f(x, t) = I(x(t), t) = I(x(t+1), t+1); \frac{\partial f(x)}{\partial t} = 0 \tag{17}$$

Temporal Persistence The motion of the moving pixel should be slow such that it is tracked properly. It should not move fast.

$$\frac{\partial I}{\partial x}|_t \left(\frac{\partial x}{\partial t}\right) + \frac{\partial I}{\partial t}|_{x(t)} = 0; I_x v + I_t = 0; v = -\frac{I_t}{I_x} \tag{18}$$

Spatial Coherence It assumes that the neighbouring pixel has similar motion as that of the moving pixel. It is assumed that all the pixels lie on the same plane. Hence, it can be said that all pixel has similar motion.

$$\begin{bmatrix} I_x(p_1) & I_y(p_1) \\ I_x(p_{25}) & I_y(p_{25}) \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = -\begin{bmatrix} I_t(p_1) \\ I_t(p_{25}) \end{bmatrix}; Ad = -B \quad (19)$$

Optical flow analysis. The output optical flow vector is obtained for each pixel from the above steps. The optical flow vectors can be defined as $p = [px, py]_i$, and $q = [qx, qy]_i$ where $i = 1, 2, \ldots, n$, p, and q are the position vectors in both frames. The velocity of the fire blob can be considered as the average velocity vector of each fire pixel and this is given by Eq. (20) as below:

$$V_{\text{avg}} = \frac{1}{n}\sum_{i=1}^{n}\sqrt{\left(p_{yi} - q_{yi}\right)^2 + \left(p_{xi} - q_{xi}\right)^2} \quad (20)$$

4 Experimental Results

The proposed system is evaluated with the help of various datasets. The dataset include fire and non-fire videos taken at different time (day, evening and night) and of different fire size (small, medium and large). The videos are of different format

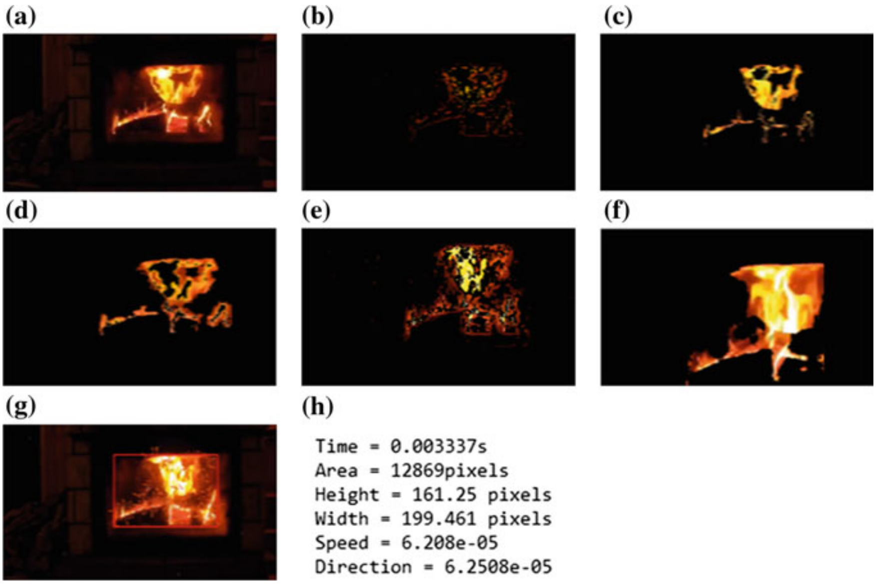


Fig. 2 **a** Input frame, **b** segmentation using RGB color model, **c** segmentation using YCbCr color model, **d** segmentation using $L^*a^*b^*$ color model, **e** dilation of color segmented image, **f** flicker image, **g** optical flow for fire detected regions, **h** fire analysis using optical flow and blob analysis

like.mp4 and .avi with varied length ranging from 2 to 14 s. Different resolution videos ranging from 240 * 320 to 720 * 1280 are considered. The proposed algorithm is implemented on Intel Core i7-4510U CPU @2.00 GHz with MATLAB R2016a. The intermediate results of the experiment are shown in Fig. 2.

The parameters used for evaluation are Accuracy and *F* measure. These can be calculated as discussed in Eqs. (21) and (22) referred from [13].

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FN} + \text{FP}} \tag{21}$$

$$F\ \text{measure} = \frac{2 * \text{Recall} * \text{Precision}}{(\text{Recall} + \text{Precision})} \tag{22}$$

TP Fire image predicted as fire image
TN Non-fire image predicted as non-fire image
FP Non-fire image predicted as fire image
FN Fire image is predicted as non-fire image.

Performance of the system is shown in Table 1. The average accuracy for the proposed fire detection system is 80.98%. The average F measure value is 89.95%.

5 Conclusion

This work proposes an early vision fire detection system that uses fire characteristics like color, area, intensity flicker, and motion for detection of fire. The fire is analyzed using optical flow and blob analysis. Blob analysis is to obtain details like location of fire, height, width while Lucas–Kanade optical flow measures the average velocity and direction of the fire. Experimental results show that the proposed work gives better results for fire videos. The average accuracy of the system is around 80.98%. Non-fire videos have less accuracy and enhancements needs to be done to improve upon the accuracy of non-fire videos. Videos with fire of very high intensity having color almost similar to white is not detected properly so enhancements for the same can be made in future to improve the false negatives of the system.

Table 1 Analysis of proposed visual fire detection system

Video	Duration (s)	No. of frames	No. of fire frames	No. of non-fire frames	TP	TN	FP	FN	Accuracy	F measure
video1	5.77	144	144	0	119	0	0	25	0.8264	0.9049
video2	3.00	61	61	0	45	0	0	16	0.826	0.8491
video3	9.00	265	265	0	110	0	0	55	0.6667	0.800
video4	3.00	59	59	0	46	0	0	13	0.7797	0.8762
video5	6.00	178	0	178	0	148	30	0	0.8315	–
video6	3.85	72	72	0	50	0	0	22	0.6944	0.8197
video7	4.00	76	76	0	51	0	0	25	0.6711	0.8031
video8	4.68	93	93	0	80	0	0	13	0.8602	0.9249
video9	3.00	52	52	0	49	0	0	3	0.9423	0.9703
video10	2.60	43	0	43	0	23	20	0	0.5349	–
video11	2.50	39	39	0	36	0	0	3	0.9231	0.9600
video12	3.64	67	67	0	62	0	0	5	0.9254	0.9612
video13	4.4	86	86	0	84	0	0	2	0.9767	0.9882
video14	3.88	73	73	0	72	0	0	1	0.9863	0.9931
video15	2.80	46	46	0	43	0	0	3	0.9348	0.9663
video16	3.45	61	61	0	58	0	0	3	0.9508	0.9758
video17	3.16	55	55	0	54	0	0	1	0.9818	0.9908
video18	5.68	118	0	118	0	68	50	0	0.5763	–
video19	8.00	212	212	0	134	0	0	78	0.6321	0.7746
video20	5.13	125	125	0	122	0	0	23	0.8414	0.9139
video21	6.2	131	131	0	65	0	0	66	0.4962	0.6633
video22	7.56	165	165	0	112	0	0	53	0.6788	0.8087
video23	8.28	183	183	0	161	0	0	22	0.8798	0.9360
video24	3.64	67	67	0	64	0	0	3	0.9552	0.9771
video25	2.24	32	32	0	28	0	0	4	0.8750	0.9333
Average									0.8098	0.8995

References

1. Celik T, Demirel H (2009) Fire detection in video sequences using a generic color model. Fire Saf J 44(2):147–158
2. Zhang Z, Zhao J, Zhang D, Qu C, Ke Y, Cai B (2008) Contour based forest fire detection using FFT and wavelet. In: 2008 International conference on computer science and software engineering, vol 1, pp 760–763
3. Celik T (2010) Fast and efficient method for fire detection using image processing. ETRI J 32(6):881–890
4. Zhang Z, Shen T, Zou J (2014) An improved probabilistic approach for fire detection in videos. Fire Technol 50(3):745–752
5. Kumar TS, Gautam KS, Haritha H (2016) Debris detection and tracking system in water bodies using motion estimation technique. In: Innovations in bio-inspired computing and applications, pp 275–284

6. Toreyin BU, Dedeoglu Y, Cetin AE (2005) Flame detection in video using hidden markov models. In: IEEE international conference on image processing, vol 2, pp II–1230
7. Bohush R, Brouka N (2013) Smoke and flame detection in video sequences based on static and dynamic features. In: Signal processing: algorithms, architectures, arrangements, and applications (SPA), pp 20–25
8. Chen T-H, Wu P-H, Chiou Y-C (2004) An early fire-detection method based on image processing. In: 2004 International conference on image processing, vol 3, pp 1707–1710
9. Yuan C, Liu Z, Zhang Y (2015) UAV-based forest fire detection and tracking using image processing techniques. In: International conference on unmanned aircraft systems, pp 639–643
10. Kecheril SS, Dr. Venkataraman D, Suganthi J, Sujathan K (2012) Segmentation of lung glandular cells using multiple color spaces. Int J Comput Sci Eng Appl 2(3):147–158
11. Stadler A, Windisch T, Diepold K (2014) Comparison of intensity flickering features for video based flame detection algorithms. Fire Saf J 66:1–7
12. Rinsurongkawong S, Ekpanyapong M, Dailey MN (2012) Fire detection for early fire alarm based on optical flow video processing. In: 9th International conference on electrical engineering/electronics, computer, telecommunications and information technology, pp 1–4
13. Kim Y-H, Kim A, Jeong H-Y (2014) RGB color model based the fire detection algorithm in video sequences on wireless sensor network. Int J Distrib Sens Netw 10(4)
14. Toulouse T, Rossi L, Akhloufi M, Celik T, Maldague X (2015) Benchmarking of wildland fire colour segmentation algorithms. IET Image Process 9(12):1064–1072

Comparative Performance Analysis of Local Feature Descriptors for Biomedical Image Retrieval

Suchita Sharma and Ashutosh Aggarwal

Abstract Biomedical imaging field is growing enormously from last decade. The medical images have been used and stored continuously for diagnosis as well as research purposes. There exist several methods that tend to provide real-time retrieval of medical images from such storage repositories. Therefore, in this paper, we strive to present an exhaustive performance comparison of existing and recently published state-of-the-art local feature descriptors for retrieval of CT and MR images. All the compared methods have been tested on two standard test databases, namely, NEMA CT and NEMA MRI. Additional experiments have been conducted to analyze the noise robustness ability of all the compared approaches. Lastly, the methods are also compared in terms of their computational complexity and total CPU time taken to retrieve images corresponding to the given query image.

1 Introduction

The usage of biomedical images during diagnosis of a patient in hospitals is increasing rapidly. The biomedical images can be in the form of X-ray, computed tomography (CT), magnetic resonance imaging (MRI), etc. Due to the rapid increment in biomedical images, it is becoming a more challenging task to diagnose a patient in clinics and hospitals. Hence, there is a need to have a more accurate and effective searching, indexing and retrieving techniques for a proper diagnosis. Moreover, it cannot be achieved without a proper arrangement of the biomedical images data. To cope up with such type of situations a new technique named content-based image retrieval (CBIR) system comes into the existence. CBIR systems have benefited the

S. Sharma (✉) · A. Aggarwal
Department of Computer Science and Engineering, Thapar Institute of Engineering and Technology, Patiala, India
e-mail: ssharma_me16@thapar.edu

A. Aggarwal
e-mail: ashutosh.aggarwal@thapar.edu

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_100

collection and management of continually increasing biomedical images as well as the medicine, R&D department, and education.

CBIR systems include a process of retrieval the most relevant images from an already existing image dataset on the basis of their primary (e.g., color, texture, and shape) or semantic features. To develop robust and efficient medical image retrieval (MIR) system, extraction of appropriate visual features like shape, color, and texture plays a significant role. These features (especially textures) provide the salient information from medical images in an efficient manner. The visual feature descriptors of an image can be categorized into two types as the local feature descriptors and the global feature descriptors.

In local feature descriptors, images present in the database are divided into the blocks and binary representation of each block is obtained by exploring the neighborhood in different directions. Then these blocks are scale to calculate local pattern and by concatenation of these local patterns, a feature vector is formed to retrieve the medical images. In global feature descriptors, the entire image is considered as a single unit for the purpose of feature extraction. The description of how the CBIR systems work with different feature extraction techniques is detailed in [1–5]. As biomedical images can be in the different formats, in this paper our aim is to retrieve the CT and MRI images. The work done in the initial days for the retrieval of MRI and CT images is mentioned in [6]. The PACS (Picture archiving and communication system), IRMA (Image retrieval for medical applications), SPIRS (Spine pathology and image retrieval system, etc., are the most popular content-based medical image retrieval systems (MIR).

Initially, local binary pattern (LBP) [7] has been proposed as a local feature descriptor for the purpose of the texture synthesis. LBP achieves very appreciable results in texture synthesis because of its simple working strategy. LBP has given a new insight into the research areas like medical, motion analyzing, face detection and image browsing and retrieval, etc. Because of its uniform direction working strategy, the results expecting under the difficult lighting conditions, noise conditions, pose conditions, aging of the images, etc., are not that much satisfactory. To deal with such type of conditions LBP is further extended to its various variants for serving the different application areas. To cope up the lightning conditions Tan and Triggs introduced a new local feature extraction technique named Local ternary patterns (LTP) [8] as a three-valued code formation method. With the help of LTP, it becomes easier to analyze the different body organs under different pose and lighting conditions. In the same trend, local quantized patterns (LQP) [9] and local tetra patterns (LTerPs) [10] are introduced for the retrieval of the biomedical images. The LTerPs have the extraordinary power to explain the spatial structure of the body organs in a more accurate and detailed way in a four coded-values format. The proposed extractor has the problem of high feature vector length. To overcome the feature vector length problem, Murala and Wu further introduced a new method of feature extraction named Local co-occurrence ternary patterns (LTCoP) [11] especially to retrieve out the MRI images. Later on, the local mesh edge patterns (LMeP) [12] for the retrieval of biomedical images has been introduced by the same authors. The working strategy in local mesh edge pattern (LMeP) is the little bit different than the already existing local methods, the LMeP simply encodes the relationship among the neighbors. After that, the LMeP has been further extended to a new feature descriptor method called Local Mesh Peak Valley edge patterns (LMePVPs) [13].

All the above-mentioned methods can work only for the two-dimensional view of the input medical grayscale images. To work with the 3D view of the medical images a new method called spherical symmetric three-dimensional local ternary patterns (SS-3D-LTP) [14] has been introduced. The SS-3D-LTP have the advantage to encode the relationship between the central pixel to its neighboring pixels in a selected five directions with an aim to convert the input 2D image to the 3D plane with the help of the Gaussian Filter Bank. From the 3D view of the body organs, it has become easier for the physicians to analyze the region of disorder more quickly and in an effective way. The LQPs are further extended to local quantized extrema patterns (LQEP) [15] for the better image retrieval. Later on, Dubey et al. introduce a new feature descriptor Local wavelet pattern (LWP) [16] especially for the CT images retrieval and the method has more discriminate power as compared to the already existing ones. The problem with LWP is to have the lesser dimensionality.

To cope with the dimensionality problem, the same author further proposed the new methods named local diagonal extrema pattern (LDEP) [17], local bit-plane decoded patterns (LBDP) [18] and local bit-plane dissimilarity patterns (LBDISP) [19] especially to achieve the better performance during the retrieval of CT images. After that, Deep et al. introduces the new methods called local mesh ternary patterns (LMeTP) [20], directional local ternary quantized extrema (DLTerQEPs) [21] and local quantized extrema quinary patterns (LQEQP) [22] to extend the standard LBP and LTP functionality in a 2D image to the selected directions especially for the retrieval of MRI and CT images. DLTerQEPs uses the ternary patterns from the horizontal, vertical, diagonal, anti-diagonal ($HVDA_7$) structure of the directional local extrema pattern values of an image to encode more spatial structure information for the better retrieval. LMeTP simply uses the ternary patterns from the mesh patterns of an image to encode the better spatial structural information of the medical images. In the same trend, Verma and Raman have been introduced a new method named local tri-diagonal patterns (LTDP) [23] to achieve the better retrieval performance rate over the LBP and all its variants by encoding all the possible information associated with a pixel in an image. LTDP has the capabilities to extract more accurate and systematic features from an input image in the comparison to the existing methods. Although the LTDP has been proved as a better feature descriptor, sometimes results may lead to the loss of discrimination information. To have a control over the loss of discrimination information the new method named local directional gradient pattern (LDGP) [24] has been introduced. In the same trend, a new feature descriptor method Local neighborhood Intensity Pattern (LNIP) [25] has been proposed. LNIP works on the idea to calculate the intensity difference between a particular pixel's adjacent neighbors. Recently, a new local descriptor method Local neighborhood difference pattern (LNDP) [26] by Verma and Raman has been introduced to find out the mutual relationship of all the neighboring pixels in a binary pattern with respect to each pixel in the image. The LNDP has the complementary results over LBP and all its variants.

The objective of the review presented in this paper is to identify those methods which are effective as well as efficient in accurate retrieval of different biomedical images.

The rest of the paper is organized as, Sect. 2 presents a brief overview of LBP. Section 3 presents the experimental framework used for testing the retrieval performances of all the compared approaches. Section 4 presents the computational complexity associated with all the compared approaches. Last, the paper is concluded in Sect. 5.

2 Existing Local Feature Extraction Methods

2.1 Local Binary Patterns

The local binary pattern (LBP) was proposed by Ojala et al. for the purpose of texture classification. LBP had success in terms of speed and there was no need to tune any parameter. For a given input grey scale image, the LBP pattern is calculated by constructing a local window around each pixel of the image to encode the intensity difference relationship between a particular central pixel (I_c) to all its neighboring pixels intensity values (I_i). The number of neighboring pixels is depends on the value of radius R. If $R = 1$ means that the eight neighbors will be lying on the circumference of the circle around the central pixel. The value obtained from the intensity difference are replaced with the binary number either 0 or 1. Thus, a binary code of a particular length is obtained for the selected central pixel, to get a LBP map the weights are assigned to the obtained binary values as described in Eq. (1)

$$\mathrm{LBP}_{N,R} = \sum_{i=1}^{N} 2^{i-1} \times D_1(I_i, I_c) \tag{1}$$

To construct a feature vector for the entire image a histogram concatenation is done for all possible obtained LBP maps as described in Eq. (3)

$$\mathrm{Hist\_LBP} = \sum_{i=1}^{X}\sum_{J=1}^{Y} D_2(\mathrm{LBP}(I_i, I_c), L); L \in \left[0, \left(2^N - 1\right)\right] \tag{2}$$

$$D_2(a_1, a_2) = \begin{cases} 1, & a_1 = a_2 \\ 0, & \text{otherwise} \end{cases} \tag{3}$$

2.2 Local Ternary Patterns

Tan and Triggs extended the LBP from the two-valued code to three-valued code. The mathematical equation to represent the LTP is as

$$\vec{F}(x, g_c, t) = \begin{cases} +1, & x \geq g_c + t \\ 0, & |x - g_c| < t \\ -1, & x \leq g_c - t \end{cases} \tag{4}$$

where t is a threshold value, values $g_c - t$ and $g_c + t$ are considered to be as the low and high range of values. The values lie in the lower range of the low are replaced as -1, values in the upper range to the high value replaced by +1 and the value lies between the high and low is considered to be as the 0. In LTP the threshold value is a user defined value. Hence, the obtained LTP values are more resistant to noise and no more effect of the invariance to the gray-level transformation. To construct a feature vector for the entire image a histogram concatenation is done for all possible obtained LTP maps as described earlier in LBP.

The readers are requested to refer to cited references in the introduction section to read about the other local feature descriptors compared in this paper.

3 Experiments and Results

In this section, we present the computational framework employed to test the performance of the existing local feature descriptors for biomedical image retrieval. We have considered the following mentioned methods in our comparative analysis. All the methods have been implemented in MATLAB 2013.

S. No.	Abbreviation	Method name	Year
1	LBP	Local Binary Pattern	2002
2	LTP	Local Ternary Pattern	2009
3	LQryP	Local Quinary Patterns	2010
4	LTrP	Local Tetra Patterns	2012
5	LTCoP	Local Ternary Co-occurrence Patterns	2013
6	LMeP	Local Mesh Patterns	2014
7	LMePVEP	Local Mesh Peak Valley Edge Patterns	2014
8	LDEP	Local Diagonal Extrema Pattern	2015
9	LWP	Local Wavelet Pattern	2015

(continued)

(continued)

S. No.	Abbreviation	Method name	Year
10	LQEP	Local Quantized Extrema Patterns	2015
11	SS-3D-LTP	Spherical Symmetric 3D Local Ternary Patterns	2015
12	LBDISP	Local Bit-Plane Dissimilarity Pattern	2016
13	LBDP	Local Bit-Plane Decoded Pattern	2016
14	DLTerQEP	Directional Local Ternary Quantized Extrema Pattern	2016
15	LMeTP	Local Mesh Ternary Patterns	2016
16	LNIP	Local Neighborhood Intensity Pattern	2017
17	LTDP	Local Tri-directional Patterns	2017
18	LNDP	Local Neighborhood Difference Pattern	2017
19	LDGP	Local Directional Gradient Pattern	2017
20	LQEQryP	Local Quantized Extrema Quinary Pattern	2017

In our experiments, we use Extended Canberra similarity measure to compare the feature vectors of the query image with that of database images. The mathematical expression for the Extended Canberra distance is as

$$D_{\text{ECD}}(t, Q) = \sum_{\tau=1}^{\dim} \frac{\left|F^Q(\tau) - F^t(\tau)\right|}{\left(F^Q(\tau) + \mu^Q\right) + \left(F^t(\tau) + \mu^t\right)},$$

$$\mu^Q = \frac{1}{\dim} \sum_{\tau=1}^{\dim} F^Q(\tau)$$

$$\mu^t = \frac{1}{\dim} \sum_{\tau=1}^{\dim} F^t(\tau) \tag{5}$$

dim represents the dimension of the feature vectors, $F^Q(\tau)$ and $F^t(\tau)$ is the τth element of the feature vectors of query image Q and database image t, respectively. In our experiments, every image in the database is used as a query image. The average precision rate (APR) average retrieval rate (ARR), F_score and mean average precision (mAP) are used as the performance measures for each retrieval method.

3.1 Dataset Used

In this paper two datasets namely NEMA CT [27] and NEMA MRI [28] consisting of CT and MR images of various body organs are used for the experimental work. The NEMA CT consisting of 600 CT images categorized it into 10 categories having

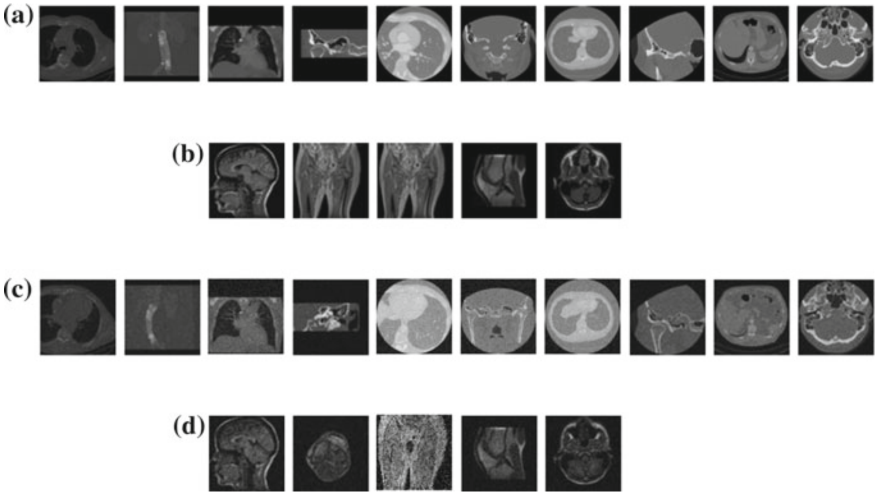


Fig. 1 Sample images from **a** NEMA CT database, **b** NEMA MRI database, **c** noisy NEMA CT database and **d** noisy NEMA MRI database

54, 70, 66, 50, 15, 60, 52, 104, 60, and 69 images. The second dataset is NEMA MRI containing 372 MR images and categorized it into 5 categories having 72, 100, 76, 59, and 65 images. Figure 1a, b shows sample image from each category of NEMA CT and NEMA MRI datasets, respectively. Besides experimenting with the above-mentioned databases, additional experiments have been performed to test the effectiveness and superiority of all the compared methods under noisy conditions. In this process, a new copy of each of the above-mentioned databases has been created consisting of noisy images. The noisy images are obtained by corrupting the original noise-free images with additive white Gaussian noise with zero mean and $\sigma \in [5, 50]$. Figure 1c, d shows sample noisy images.

3.2 Experimental Results

This section provides several experiments to demonstrate the effectiveness of existing local feature extraction methods and for all these approaches, we have taken a window size of 3×3 with $(P, R) = (8, 1)$.

3.2.1 Results on Noise-Free NEMA CT Database

Table 1 compares the retrieval performance of the all existing local feature extraction methods in terms of ARR, ARP, F_score, and mAP. The experiments have been performed to obtain the best 100 matches from the NEMA CT database. From Table 1,

Table 1 Comparison of performances of all existing local feature extraction techniques on the bases of ARR, APR, F_score and mAP for best $\eta = 100$ on noise-free NEMA CT database and all the mentioned values are in percentages (%)

	ARR	APR	F_score	mAP
LMePTerP	96.36	64.98	77.62	98.55
DLTerQEP	96.45	64.92	77.61	99.00
LQEP	96.43	64.70	77.44	98.96
LQEQerP	95.83	64.65	77.21	98.56
LMeP	96.19	64.64	77.32	98.12
LNDP	96.42	64.62	77.38	98.07
SS-3D-LTP	95.92	64.62	77.22	98.43
LTP	95.99	64.52	77.17	98.83
LTDP	96.18	64.50	77.21	98.50
LTrP	95.92	64.26	76.96	97.92
LQP	95.27	64.00	76.56	98.48
LDGP	95.52	63.98	76.63	97.27
LNIP	95.45	63.93	76.57	97.68
LDEP	95.02	63.72	76.28	97.23
LTCoP	95.47	63.55	76.31	98.36
LMePVEP	93.51	62.91	75.22	97.43
LBP	94.19	62.25	74.96	97.85
LWP	80.37	52.76	63.70	89.89
LBDP	76.91	50.06	60.64	84.45
LBDISP	71.14	45.20	55.28	82.41

it can be observed that the results obtained from DLTerQEP method are better than those obtained from the other existing methods. The compared method has the ARR, APR, F_score and mAP retrieval rates of 96.45% (ARR), 64.92% (APR), 77.61% (F_score), and 99.00% (mAP), with maximum gain of 0.79% (ARR), 3.11% (APR), 2.45% (F_score), and 0.12% (mAP) over other methods. The APR graph for all the compared methods has been shown in Fig. 2. The graph demonstrates the variation in the retrieval rates of all the compared methods with the number of top matches.

3.2.2 Results on Noise-Free NEMA MRI Database

Figure 3 shows the graphical comparison of the APR rate of all existing local feature extraction methods. The retrieval rates of all the compared methods for top 100 matches are shown in Table 2. The methods LBP, LTP, LTrP, LTCoP, LNIP, LTDP, LMeP, LMePVEP, LNDP, LDGP, LQEP, and DLTerQEP gave similar and maximum retrieval performance of 100% (ARR), 77.06% (APR), 87.04% (F_score), and 100% (mAP). Among all the compared methods, the retrieval performance of LWP has been

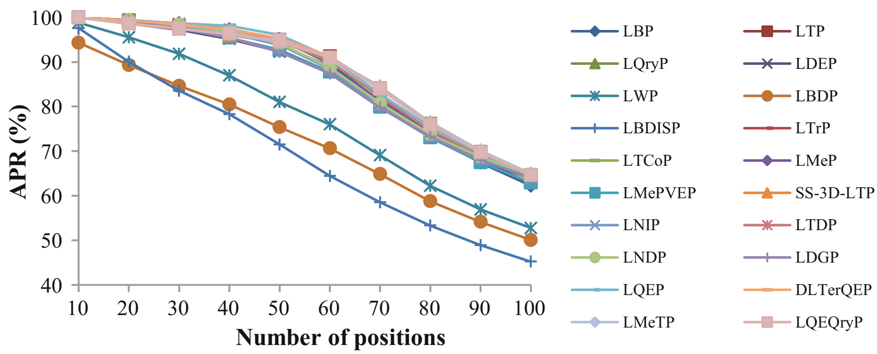


Fig. 2 Comparison between all existing local techniques on the bases of APR on noise-free NEMA CT database

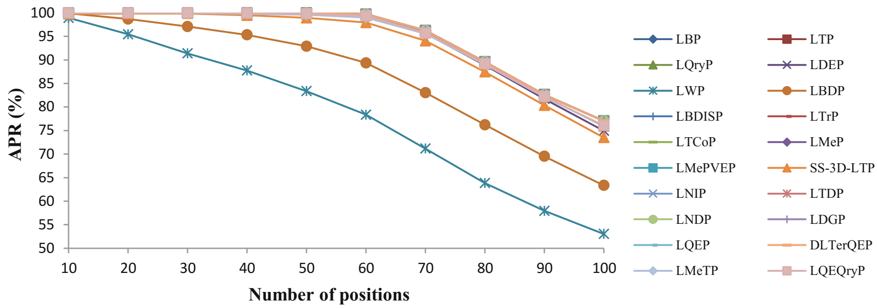


Fig. 3 Comparison between all the existing techniques on the bases of APR on noise-free NEMA MRI database

observed to be the lowest with the loss in retrieval rates of 28.49% (ARR), 27.26% (APR), 28.16% (F_score), and 13.32% (mAP).

3.2.3 Results on Noisy Databases

Tables 3 and 4 summarizes the retrieval performance of all the compared methods for top 100 matches and Figs. 4 and 5 shows the APR retrieval rates of all existing local feature extraction methods tested over noisy NEMA CT and noisy NEMA MRI databases, respectively. The method LBDP obtained the highest retrieval rates of 47.11% (ARR), 32.57% (APR), 38.51% (F_score), and 35.83% (mAP) on noisy NEMA CT database and 27.01% (ARR), 15.94% (APR), 20.13% (F_score), and 40.69% (mAP) and 60.54% (ARR), 46.20% (APR), 52.40% (F_score), and 50.17% (mAP) on noisy NEMA MRI database. From the results, it is observed that the method LBDP is highly robust to noise among all the compared methods and is the most suitable local feature extraction method for the retrieval of noisy CT and MR images.

Table 2 Comparison of performances of all existing local feature extraction techniques on the bases of ARR, APR, F_score and mAP for best $\eta = 100$ on noise-free NEMA MRI database and all the mentioned values are in percentages (%)

	ARR	APR	F_score	mAP
LBP	100.00	77.06	87.04	100.00
LTP	100.00	77.06	87.04	100.00
LTrP	100.00	77.06	87.04	100.00
LTCoP	100.00	77.06	87.04	100.00
LMeP	100.00	77.06	87.04	100.00
LMePVEP	100.00	77.06	87.04	100.00
LNIP	100.00	77.06	87.04	100.00
LTDP	100.00	77.06	87.04	100.00
LNDP	100.00	77.06	87.04	100.00
LDGP	100.00	77.06	87.04	100.00
LQEP	100.00	77.06	87.04	100.00
DLTerQEP	100.00	77.06	87.04	100.00
LBDISP	99.99	77.05	87.03	99.99
LQEQerP	98.83	76.06	85.96	99.86
LMePTerP	98.80	76.06	85.95	99.77
LQP	98.79	75.95	85.88	99.80
LDEP	97.90	74.96	84.91	99.63
SS-3D-LTP	96.22	73.51	83.34	98.87
LBDP	83.80	63.36	72.16	93.93
LWP	71.51	53.04	60.91	86.68

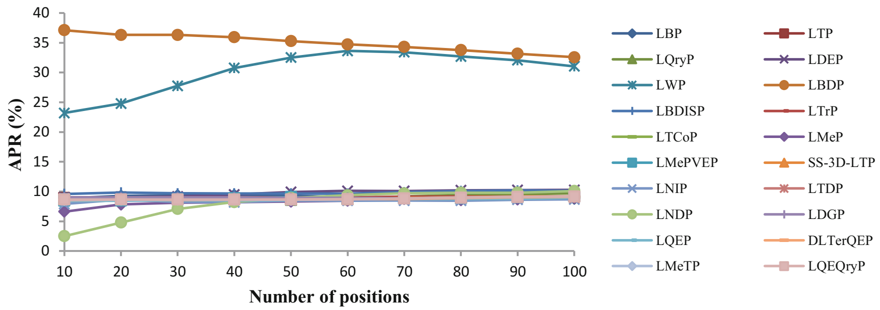


Fig. 4 Comparison between all the existing techniques on the bases of APR on noisy NEMA CT database

4 Computational Efficiency

Biomedical image retrieval requires methods which are effective as well as computationally efficient. Such methods should provide real-time response, i.e., the time

Table 3 Comparison of performances of all existing local feature extraction techniques on the bases of ARR, APR, F_score and mAP for best $\eta = 100$ on noisy NEMA CT database and all the mentioned values are in percentages (%)

	ARR	APR	F_score	mAP
LBDP	47.11	32.57	38.51	35.83
LWP	46.51	31.03	37.23	31.00
LDEP	18.39	10.27	13.18	9.48
LBDISP	19.05	10.11	13.21	9.71
LNDP	16.42	10.09	12.50	7.88
LQP	18.45	9.73	12.74	8.83
SS-3D-LTP	16.75	9.24	11.91	8.77
LTP	16.80	9.18	11.87	8.87
LMePVEP	16.81	9.18	11.87	8.69
LMePTerP	16.67	9.18	11.84	8.61
LQEQerP	16.67	9.18	11.84	8.68
LTrP	17.03	9.16	11.91	9.03
LTCoP	16.77	9.14	11.83	8.72
DLTerQEP	16.68	9.10	11.77	8.66
LDGP	17.28	9.05	11.88	8.99
LTDP	16.59	8.97	11.64	8.70
LQEP	16.55	8.93	11.60	8.52
LMeP	16.66	8.91	11.61	7.67
LBP	16.59	8.79	11.49	9.22
LNIP	16.24	8.67	11.31	8.01

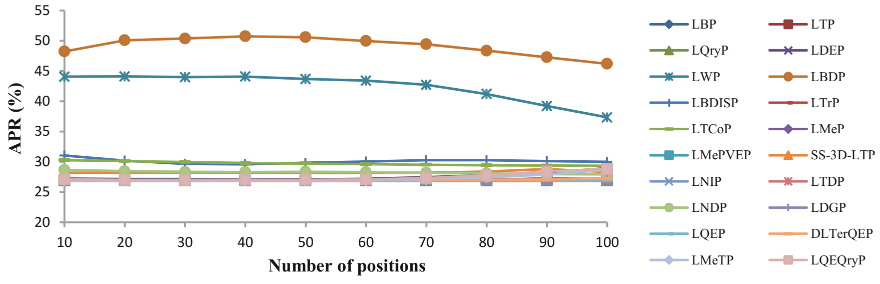


Fig. 5 Comparison between all the existing techniques on the bases of APR on noisy NEMA MRI Database

associated with the retrieval of medical images should be very less. The computational time for any retrieval system depends on two stages: (1) Feature Extraction Stage and (2) Retrieval (Classification) Stage. Table 5 shows the comparison of the feature vector length as well as the total CPU time of all the compared approaches for a given query image. The method LQEQerP has the highest number of features, i.e., 16,384 features. From table, it can be observed that the total CPU time for the

Table 4 Comparison of performances of all existing local feature extraction techniques on the bases of ARR, APR, F_score and mAP for best $\eta = 100$ on noisy NEMA MRI database and all the mentioned values are in percentages (%)

	ARR	APR	F_score	mAP
LBDP	60.54	46.20	52.40	50.17
LWP	46.34	37.33	41.35	42.04
LBDISP	34.91	29.98	32.26	30.05
LTCoP	30.36	29.33	29.84	29.92
LDEP	29.84	28.96	29.39	27.08
LQEQerP	30.04	28.83	29.42	27.04
LMePTerP	29.45	28.69	29.06	26.90
SS-3D-LTP	29.56	28.28	28.91	28.06
LNDP	29.90	27.92	28.88	28.14
LMeP	27.37	27.20	27.29	26.92
DLTerQEP	27.33	27.20	27.26	26.92
LBP	27.28	27.14	27.21	27.14
LQP	28.36	27.05	27.69	26.63
LDGP	26.96	26.94	26.95	26.95
LMePVEP	26.90	26.89	26.90	26.88
LTP	26.89	26.89	26.89	26.88
LNIP	26.89	26.88	26.88	26.88
LTDP	26.89	26.88	26.89	26.88
LTrP	26.88	26.88	26.88	26.88
LQEP	26.88	26.88	26.88	26.88

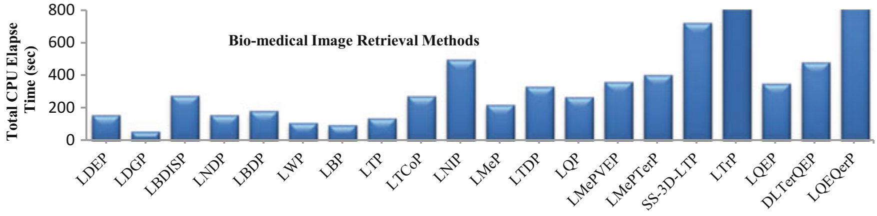


Fig. 6 Total CPU Elapse time (s) comparison of all the existing biomedical retrieval approaches over NEMA CT database

LDGP method is lowest among all other compared (existing) methods. Figures 6 and 7 show the total CPU elapse time (s) of all the compared approaches for the NEMA CT and NEMA MRI databases.

Table 5 CPU elapse time (s) taken for extracting the features and retrieved out the images from NEMA CT and NEMA MRI databases images using already existing local feature extraction approaches

Feature descriptors	No. of features	Feature extraction time (*A*)		Retrieval time (*B*)		Total CPU time (*A* + *B*)	
		NEMA CT	NEMA MRI	NEMA CT	NEMA MRI	NEMA CT	NEMA MRI
LDEP	24	156.13	24.20	1.86	1.42	157.99	25.62
LDGP	64	51.38	7.96	4.33	3.07	55.71	11.03
LBDISP	256	265.07	41.09	11.51	7.25	276.58	48.34
LNDP	256	146.43	22.70	11.51	7.25	157.94	29.95
LBDP	256	171.88	26.64	11.51	7.25	183.39	33.89
LWP	256	98.25	15.23	11.51	7.25	109.76	22.48
LBP	256	85.03	13.18	11.51	7.25	96.54	20.43
LTP	512	117.54	18.22	20.54	12.26	138.08	30.48
LTCoP	512	252.74	39.17	20.54	12.26	273.28	51.43
LNIP	512	477.18	73.96	20.54	12.26	497.72	86.22
LMeP	768	190.27	29.49	30.21	16.41	220.48	45.9
LTDP	768	302.54	46.89	30.21	16.41	332.75	63.3
LQP	1024	226.25	35.07	41.68	21.98	267.93	57.05
LMePVEP	1024	318.63	49.39	41.68	21.98	360.31	71.37
LMePTerP	1536	343.45	53.23	60.46	31.38	403.91	84.61
SS-3D-LTP	2560	629.75	97.61	93.82	43.03	723.57	140.64
LTrP	3328	7848.78	1216.56	121.70	51.80	7970.48	1268.36
LQEP	4096	202.86	31.44	147.91	63.01	350.77	94.45
DLTerQEP	8192	210.02	32.55	271.52	118.62	481.54	151.17
LQEQerP	16,384	385.55	59.76	510.18	252.26	895.73	312.02

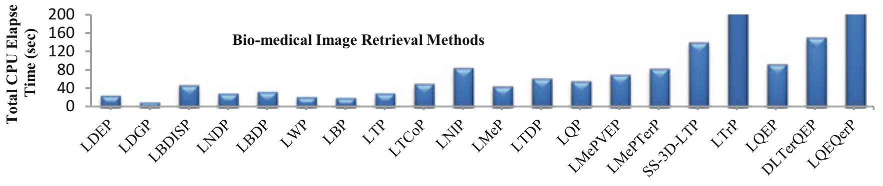


Fig. 7 Total CPU elapse time (s) comparison of all the existing biomedical retrieval approaches over NEMA MRI database

5 Conclusion

In this paper, a comparison between all the existing local biomedical image retrieval approach for CT and MR images has been presented. Experiments have been performed over both the noisy as well as noise-free images of two standard databases namely NEMA CT and NEMA MRI to check the performance comparison of existing and recently published state-of-the-art local feature descriptors. After the thorough analysis of the obtained results, it was observed that the LDGP approach is most suitable for retrieval of noise-free images. LDGP is found to be the most balanced approach in terms of its retrieval performance and the computational time requirements. However, in the case of noisy datasets, LBDP and LWP approach achieved maximum retrieval rates. The higher retrieval rates demonstrate their superior noise robustness capabilities in comparison to other compared approaches.

References

1. Rui Y, Huang TS (1999) Image retrieval: current techniques, promising directions and open issues. J Vis Commun Image Represent 10:39–62
2. Smeulders AWM, Worring M, Gupta SA, Jain R (2000) Content-based image retrieval at the end of the early years. IEEE Trans Pattern Anal Mach Intell 22:1349–1380
3. Kokare M, Chatterji BN, Biswas PK (2002) A survey on current content-based image retrieval methods. IETE J Res 48:261–271
4. Liu Y, Zhang D, Lu G, Ma WY (2007) A survey of content-based image retrieval with high-level semantics. Pattern Recogn 40:262–282
5. Muller H, Michoux N, Bandon D, Geisbuhler A (2004) A review of content-based image retrieval systems in medical applications—clinical benefits and future directions. Int J Med Inform 73:1–23
6. Manjunath KN, Renuka A, Niranjan UC (2007) Linear models of the cumulative distribution function for content-based medical image retrieval. J Med Syst 31:433–443
7. Ojala T, Pietikäinen M, Harwood D (1996) A comparative study of texture measures with classification based on featured distributions. Pattern Recogn 29(1):51–59
8. Tan X, Triggs B (2010) Enhanced local texture feature sets for face recognition under difficult lighting conditions. IEEE Trans Image Process 19(6):1635–1650
9. Hussain ul S, Triggs B (2012) Visual recognition using local quantized patterns. ECCV 2012. Part II, LNCS 7573, Italy, pp 716–729
10. Murala S, Maheshwari RP, Balasubramanian R (2012) Local tetra patterns: a new feature descriptor for content-based image retrieval. IEEE Trans Image Process 21(5):2874–2886
11. Murala S, Wu QMJ (2013) Local ternary co-occurrence patterns: a new feature descriptor for MRI and CT image retrieval. Neurocomputing 119:399–412
12. Murala S, Wu QMJ (2014) Local mesh patterns versus local binary patterns: biomedical image indexing and retrieval. IEEE J Biomed Health Inform 18(3):929–938
13. Murala S, Wu QMJ (2014) MRI and CT image indexing and retrieval using local mesh peak valley edge patterns. Sig Process Image Commun 29(3):400–409
14. Murala S, Wu QMJ (2015) Spherical symmetric 3D local ternary patterns for natural, texture and biomedical image indexing and retrieval. Neurocomputing 149:1502–1514
15. Rao LK, Rao DV (2015) Local quantized extrema patterns for content-based natural and texture image retrieval. Human-Centric Comput Inf Sci 5(1):26
16. Dubey SR, Singh SK, Singh RK (2015) Local wavelet pattern: a new feature descriptor for image retrieval in medical CT databases. IEEE Trans Image Process 24(12):5892–5903
17. Dubey SR, Singh SK, Singh RK (2015) Local diagonal extrema pattern: a new and efficient feature descriptor for CT image retrieval. IEEE Signal Process Lett 22(9):1215–1219
18. Dubey SR, Singh SK, Singh RK (2016) Local bit-plane decoded pattern: a novel feature descriptor for biomedical image retrieval. IEEE J Biomed Health Inform 20(4):1139–1147

19. Dubey SR, Singh SK, Singh RK (2016) Novel local bit-plane dissimilarity pattern for computed tomography image retrieval. Electron Lett 52(15):1290–1292
20. Deep G, Kaur L, Gupta S (2016) Directional local ternary quantized extrema pattern: a new descriptor for biomedical image indexing and retrieval. Eng Sci Technol Int J 19(4):1895–1909
21. Deep G, Kaur L, Gupta S (2016) Local mesh ternary patterns: a new descriptor for MRI and CT biomedical image indexing and retrieval. Comput Methods Biomech Biomed Eng: Imaging Vis 6(2):155–169
22. Verma M, Raman B (2016) Local tri-directional patterns: a new texture feature descriptor for image retrieval. Digit Signal Proc 51:62–72
23. Deep G, Kaur L, Gupta S (2017) Local quantized extrema quinary pattern: a new descriptor for biomedical image indexing and retrieval. Comput Methods Biomech Biomed Eng: Imaging Vis 1–17
24. Chakraborty S, Singh SK, Chakraborty P (2017) Local directional gradient pattern: a local descriptor for face recognition. Multimedia Tools Appl 76(1):1201–1216
25. Banerjee P, Bhunia AK, Bhattacharyya A, Roy PP, Murala S (2017) Local neighborhood intensity pattern. A new texture feature descriptor for image retrieval. arXiv preprint arXiv: 1709.02463
26. Verma M, Raman B (2017) Local neighborhood difference pattern. A new feature descriptor for natural and texture image retrieval. Multimedia Tools Appl 1–24
27. NEMA CT Image Database [Online] ftp://medical.nema.org/medical/Dicom/Multiframe/CT
28. NEMA MR Image Database [Online] ftp://medical.nema.org/medical/Dicom/Multiframe/MR

Detection of Liver Tumor Using Gradient Vector Flow Algorithm

Jisha Baby, T. Rajalakshmi and U. Snekhalatha

Abstract Liver tumor also known as the hepatic tumor is a type of growth found in or on the liver. Identifying the tumor location can be a tedious, error-prone and need an experts study to identify it. This paper presents a segmentation technique to segment the liver tumor using Gradient Vector Flow (GVF) snakes algorithm. To initiate snakes algorithm the images need to be insensitive to noise, Wiener Filter is proposed to remove the noise. The GVF snake starts its process by initially extending it to create an initial boundary. The GVF forces are calculated and help in driving the algorithm to stretch and bend the initial contour towards the region of interest due to the difference in intensity. The images were classified into tumor and non-tumor categories by Artificial Neural Network Classifier depending on the features extracted which showed notable dissimilarity between normal and abnormal images.

1 Introduction

There are many types of liver malfunction found in day to day life one among them being liver tumors. Liver tumor is also known as the hepatic tumors are growths found in or around the area of liver. There are various types of liver tumors which can be formed as liver is made up of different cell types. These growths found in the liver can be benign or malignant which are cancerous growths. It can be found using various liver function test which helps in finding out the initial liver problem. Liver tumors can be found only using any of the imaging techniques. The radiologists carefully study the liver image from the scans to find the accurate location of the tumor, if present. Since these are done manually it takes time to find the location of tumor, they are also more prone to errors and requires an experts study to find

J. Baby (✉) · T. Rajalakshmi · U. Snekhalatha
Department of Biomedical Engineering, SRM Institute of Science and Technology, Kattankulathur, Chennai 603203, Tamil Nadu, India
e-mail: jisha_babykutty@srmuniv.edu.in

T. Rajalakshmi
e-mail: abirajalakshmix@gmail.com

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_101

the exact location. The study done in this paper is about segmenting the region of interest from the CT imaging techniques using the Gradient Vector Flow (GVF) Snakes Algorithm.

A snakes algorithm is a type of Active contour which are used to initiate a curve in the image to detect boundaries of an object in facial recognition [1, 2], medical image processing [3–5], image segmentation [6] and tracking of objects [7, 8]. An initial contour is set on the object to create an initial boundary which goes through various iteration allowing the contour to converge into the true boundaries of the object. The algorithm can be used to find an organs outline from any imaging techniques or to identify a part form the unit being manufactured. From the instigation of Snakes algorithm by Kass et al. [9] in the year of 1988 there are two problems which could not be solved. One of the problems was that the initial contour could not move to the objects which were situated far away and the other was that it was not able to get into any concavities present in boundary. The introduction of GVF snake has proven to be much better than the previous methods and also has a much faster rate in segmenting the desired part. Gradient Vector Flow snakes algorithm was used by various researchers to segment as well as to find the location of the desired part or region of interest.

2 Related Work

Bahreini et al. in 2010 used GVF algorithm to segment the lesions present in the Breast from the MRI Images [10]. The study was conducted on 52 female who had different types of lesions on the breast. In these 52 females who had lesions, 33 were malignant and 19 were benign lesions. The study included selecting the required region, segmenting the region of interest using GVF algorithm and being verified by the manual segmentation done by the radiologist's. Gradient Vector Flow Snake segmentation technique was able to segment 97% of the malignant and 89.5% of benign lesion borders with a threshold overlap of 0.6 thus helping to make accurate segmentation on the borders of breast lesions.

Díaz-Parra et al. [11] used Gradient vector flow algorithm in 2014 on computed tomography images for a fully automated spinal canal segmentation for radiation therapy. This paper proposed a fully automated Gradient Vector Flow-based snake algorithm which is used for segmenting the canal of a spine. This method was conducted on three patients with the help of Dice coefficient and the results were inferred as 79.5, 81.8 and 83.7%.

Gradient Vector Flow Algorithm was used in medical image segmentation in 2012 by Yuhua et al. [12]. This paper proposed Sigmoid Gradient Vector Flow (SGVF) algorithm which was used to improve the performance in the contour. The proposed method can avoid leakage from the weak edges as the external energy of the algorithm is callous to the noises. The algorithm was used on both synthetic and real images of an ultrasound and MRI imaging techniques which removed the noises and also extracted the edges which were weak. The results of this method presented that the algorithm lead to much more precise segmentation.

Qiongfei et al. [13] used Gradient Vector Flow Model for infrared image segmentation in 2015. This paper used the extended neighborhood along with the Gradient Vector Flow algorithm which analyses a pixel of an image and then checks the neighboring pixel. If any of the neighboring pixel is bigger than the previous then it replaces the original pixel. This calculation technique uses lesser number of iterations and the results acquired is also better. The new calculation method only needs less iteration numbers and also obtained better effects.

Mahmoud MKA et al. [14] used Gradient Vector Flow snakes algorithm to segment skin cancer images in 2011. In this paper images of skin cancer is segmented using the GVF snakes algorithm. For the hair on the skin to be removed and the noises to be calloused an adaptive filter consisting of Wiener and Median filter is used. The boundary of the skin cancer can be tracked even with the presence of any other objects near the region of interest by extending the algorithm to one direction. This research was conducted on eight skin cancer patients and the algorithm was considered to be effective.

3 Proposed Work

The methodology of liver tumor segmentation goes forward in various stages using Matlab software. As per the diagram shown in Fig. 1, Stage one consists of the original input image of the liver which has been acquired from the databases [15–17].

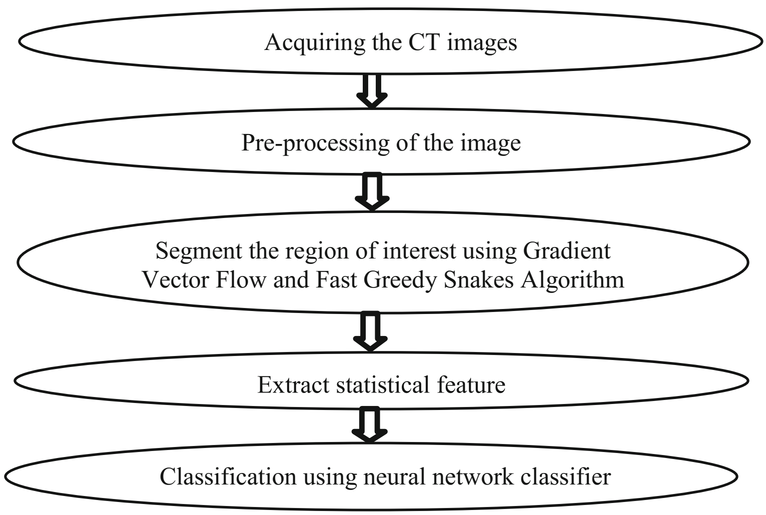


Fig. 1 Methodology for liver tumor segmentation

Stage two consists of preprocessing the images which includes converting the color image into greyscale image and also to resize the image of 0–255 size, as 0–125 and 135–255 can identify the difference between its pixel but that of in the between has minute changes which can be only detected if its resized to that scale. In stage three we propose use of Wiener filter, this filter is used to remove the unwanted noises present in the image. Stage three is used to initialize the Gradient vector flow GVF snakes algorithm to initialize the contour on the true image and segment it automatically. In stage four it is proposed to give the features extracted of tumor and non-tumor images as input to the Artificial Neural Network Classifier to train it to classify itself into tumor images or non-tumor images.

Preprocessing of the image removes distortions and enhances the image quality for better image and processing. The preprocessed image is then filtered and can be segmented using the proposed segmentation algorithm. The segmented image is given to the backpropagation Neural Network classifier to classify the images between tumor and non-tumor liver. The classifier used here has three different layers which are the input layer, hidden layer and the output layer which are added with weights for training purpose. These weights are modified during the training process to make an appropriate decision when it comes through similar input. The main reason for the use of this classifier was due to its ability to generate complex decision boundaries in any situation.

3.1 Preprocessing and Image Filtering

Preprocessing [15, 16] is used to improve the data which helps in subduing the difficult noises and also increases the quality of the images so that it can undergo further processing. Pre-processing of images in this section consists of resizing the image and then converting them to grayscale image using various Matlab commands.

The Wiener filter [16, 17] is used to generate an estimated signal by giving the input as the known signal and comparing it to the unknown signal by computing the unknown signal to give an output similar to the known signal. This filter can be used to remove any type of distortions present in an image and acquire only the interested signal. Some of the reasons why wiener filter was used over the other filter was because it did not require noncausal or causal filter and FIR(Finite Impulse Response) for filtering the images.

3.2 Image Segmentation

The process of slicing the images into various parts is called Image segmentation. Segmentation technique is used to gather peculiar information from a digital library or to locate objects of interest from a vast amount of objects. The method used here is Gradient Vector Flow snakes algorithm. Gradient vector flow (GVF) [18–21] snake algorithm, starts off by calculating the GVF forces. These forces help the initial contour to mold and stretch to the boundaries of the region of interest. The GVF forces are found using generalized diffusion equations to both components of the gradient of an image edge map. Energy function $L(c)$ of a GVF algorithm defined on the contour is given in Eq. (1):

$$L(c) = \text{Internal energy}(L_{\text{int}}) + \text{External energy}(L_{\text{ext}}) \tag{1}$$

$$L(c) = L_{\text{int}} + L_{\text{ext}} \tag{2}$$

where L_{int} and L_{ext} denote the internal and external energies, respectively. The initial contour converges to the contour smoothly with the help of the internal energy. The contour development is categorized using the external energy. The contour fits into the edges due to the image energy calculated. The external energy is said to be zero when the contour is on the boundary of an object as the position will not change. If there is any change in the position of the contour is due to the difference in the control points.

In this paper, the algorithm here uses GVF algorithm to segment the image by initializing an initial contour to the tumor liver CT image. The snake's algorithm consists of energy function as shown in Eq. (3)

$$L = \int_0^1 L_{\text{int}}(v(s)) + L_{\text{image}}(v(s)) + L_{\text{con}}(v(s))\mathrm{d}s \tag{3}$$

In here we use the GVF field as the constraint energy on the original snake's energy equation as given above. To get the GVF field we need to extract the edge map function from the image function. The GVF field is calculated using Eq. (4)

$$L = \iint \mu\left(ux^2 + uy^2 + vx^2 + vy^2\right) + |\nabla f|^2\left|g - \nabla f^2\right|\mathrm{d}x\mathrm{d}y \tag{4}$$

The smoothening term also known as the data term denoted as "μ" which is set as $1 * \mathrm{e}^{-7}$ adjusts the trade off between the 1st and 2nd term as the value of the data term (μ) decreases simultaneously the noise too decreases. As the GVF field is found its applied to the snakes algorithm [10–15, 22].

3.3 Feature Extraction and Classification

This phase is important as it extracts various features of the images helping them to provide information to the classifier. Feature extraction is used to find data which can be built to be informative, help in further learning, better interpretations, and get derived values necessary. The features which are selected need to be accurate and in similar form for the classification to be correct. The selected features are required to contain information's acquired from the input data in order to undergo the desired work. Following features were found from the segmented image for the purpose of classification.

Mean—It finds the basic texture feature representing the average pixel value of the image useful for delivering the background.

$$\text{Mean} = \frac{1}{\mathbf{n}} * \sum \mathbf{x} \tag{5}$$

Standard Deviation—It indicates how a histogram is spread along the image. It indicates what other pixel values belong to the part of the background

$$\text{Standard Deviation} = \sqrt{\frac{1}{N} * \sum (x - \mu)^2} \tag{6}$$

Correlation—It allows the image to compare for two different signals allowing to detect any difference in this waveforms.

$$\text{Correlation} = \frac{N \sum xy - \left(\sum x\right)\left(\sum y\right)}{\sqrt{[N \sum x^2 - \sum x^2][N \sum y^2 - \sum y^2]}} \tag{7}$$

Contrast—It is the difference between the maximum and minimum pixel intensity of an image. Viewing of the image is made easier with the help of this feature.

Kurtosis—It is interpreted in combination with the noise and resolution measurement.

$$\text{Kurtosis} = n \frac{\sum_{i=1}^{n} (x_i - x_{\text{avg}})^4}{\left(\sum_{i=1}^{n} (x_i - x_{\text{avg}})^2\right)^2} \tag{8}$$

Skewness—It helps in making a judgment about the surface of the image.

$$\text{Skewness} = \sqrt{n} \frac{\sum_{i=1}^{n} (x_i - x_{\text{avg}})^3}{\left(\sum_{i=1}^{n} (x_i - x_{\text{avg}})^2\right)^{3/2}} \tag{9}$$

Homogeneity—It is used to check if all the pixel in the image has same degree of value.

Energy—It is calculated using mean squared value of the image signal. It is a signal used to help in finding the frequency being distributed.

$$\text{Energy} = 1/2 \text{ mv}^2 + \text{mgh} \tag{10}$$

Back Propagation Artificial Neural Network Classifier is used to classify the proposed system using the features extracted. The classifier consists of three layers input, hidden and output layer which are connected to each other using weights. These weights are modified during the training process to make an appropriate decision

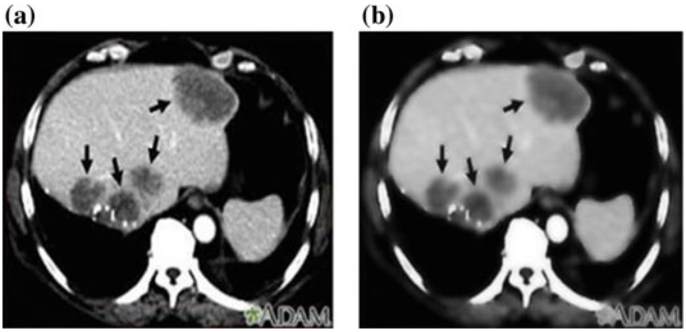


Fig. 2 **a** Input image **b** filtered image

when it comes through similar input. The main reason for the use of this classifier was due to its ability to generate complex decision boundaries in any situation. The features extracted were given as input and was trained by the classifier using the various layers present in the classifier to give the true positive, true negative, false positive, and false negative values. The images were classified as normal or abnormal CT images using this classifier. The formula for finding the confusion matrix is given by Eqs. (11)–(13):

$$\text{Accuracy} = \frac{\sum \text{True Positive} + \sum \text{False Positive}}{\sum \text{Total Population}} \tag{11}$$

$$\text{False Discovery Rate} = \frac{\sum \text{False Positive}}{\sum \text{Predicted Condition Positive}} \tag{12}$$

$$\text{Negative Predictive Value} = \frac{\sum \text{True Negative}}{\sum \text{Predicted Condition Negative}} \tag{13}$$

4 Results Analysis

The experiment was carried out on liver CT images taken from different database, on 25 abnormal image and 25 normal image. The original liver CT image as shown in Fig. 2a is provided as input. These images were preprocessed to give a resized image and then converted to grayscale image. It reduces the complexity and increases the accuracy of the applied algorithm. The Wiener filter is used to reduce the noise and reconstruct an image to perform morphological operations. Figure 2b shows the Wiener-filtered images are segmented using Gradient Vector Flow Snake Algorithm where an initial contour is initiated as shown in Fig. 4a. With respect to the initial contour, the internal and external energy is calculated and helps the snake to move towards its desired region of interest which are shown in Fig. 4b, c, d. Figure 5 shows the final segmented image (Fig. 3).

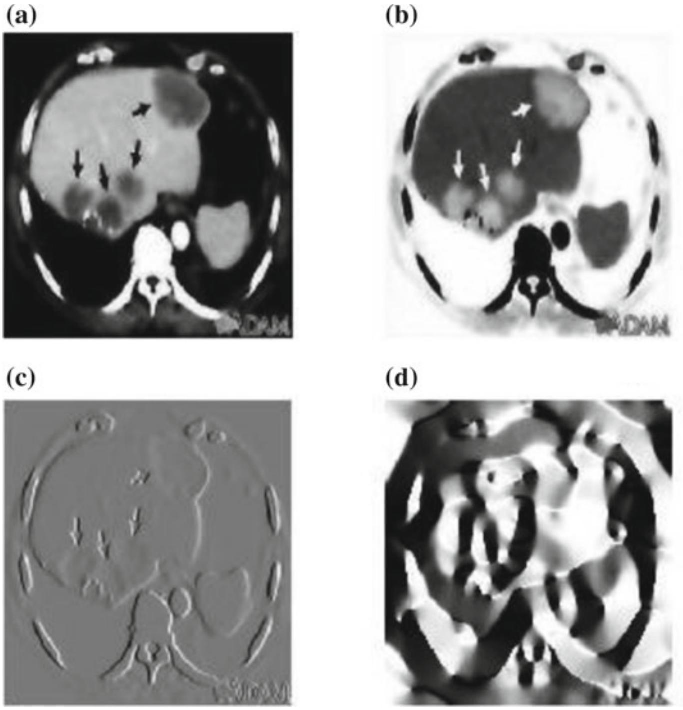
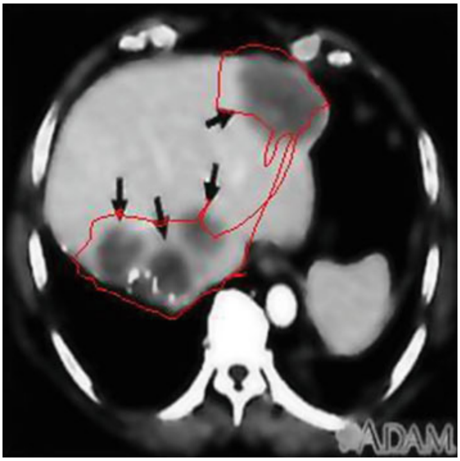


Fig. 3 Segmentation process **a** test image **b** edge map **c** edge map gradient **d** normalized GVF field

Fig. 4 Segmented image



Various statistical features calculated are tabulated in Table 1 from which it can be inferred that mean of normal CT images has higher value than the abnormal CT images, kurtosis of the abnormal images should be closer or lesser than 3 and that of

Fig. 5 Regression curve

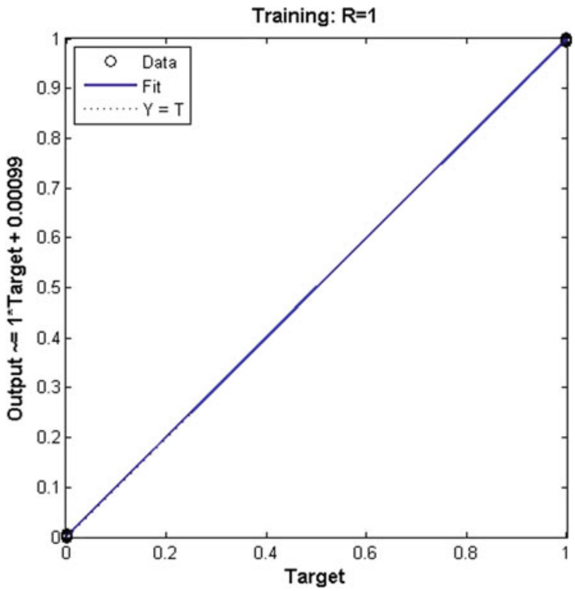


Table 1 Feature values of normal and abnormal CT image

S. No.	Feature extraction	Normal CT images	Abnormal CT images
1.	MEAN	9.508627e+01	7.670058e+01
2.	Kurtosis	4.282876e+00	3.045479e+00
3.	Skewness	3.861835e−01	7.492599e−01
4.	Correlation	9.734316e−01	9.437237e−01
5.	Homogeneity	9.331582e−01	9.921990e−01
6.	Contrast	2.003370e−01	4.368566e−01
7.	Energy	2.781193e−01	8.327415e−01
8.	Standard deviation	6.397086e+01	8.260880e+01

normal images will be greater than 3. Skewness is lesser in normal CT images than that of abnormal CT image. Correlation is slightly higher in the normal CT images. Homogeneity is higher in normal CT images. The value of contrast will be higher in abnormal CT images than the normal CT images. Energy difference is very high in abnormal CT image and the Standard deviation found is lesser in normal CT images.

The feature extracted is then given as input to the neural network classifier which classifies the image into tumor or non-tumor liver. The confusion matrix was calculated to find the accuracy of the system by calculating the true positive, negative predictive values, true negative, false discovery rate, false positive, predicted condition positive, predicted condition negative and false negatives values of the images and the accuracy was found to be 88% as shown in Table 2.

Table 2 Confusion matrix table

Total population	Condition positive	Condition negative	Accuracy = 88%
Positive predicted condition	True positive = 23	False positive = 4	False discovery rate = 14.8%
Negative predicted condition	False negative = 2	True negative = 21	Negative predictive values = 91.3%

The regression curve is used to study the linearity of the signals generated by studying the relationship between one or several signals. Regression curve is used for prediction and forecasting by learning in the Neural Networks is plotted and shown in Fig. 5. This helps in classifying the liver into normal or abnormal images.

5 Conclusion

This paper presented liver tumor image segmentation and its classification. A total of 50 images containing 25 normal and abnormal images each were used. The images were given as input and then the images were filtered using Wiener filter after going through the preprocessing. The filtered image went through the segmentation algorithm proposed in the paper where the region of interest was segmented and various features were extracted which was given as input to the neural network classifier to train itself to classify itself into tumor or non-tumor images whenever an image is given as input.

For the future work a better algorithm Fast Greedy Snakes Algorithm is used to segment the region of interest more accurately and faster than the current segmentation method.

References

1. Lam KM, Yan H (1994) Fast algorithm for locating head boundaries. J Electron Imaging 3(4):352–359
2. Waite JB, Welsh WJ (1990) Head boundary location using snakes. Br Telecom Technol J 8(3):127–135
3. McInernery T, Terzopolous D (1996) Deformable models in medical image analysis: a survey. Med Image Anal 1(2):91–108
4. Jain AK, Smith SP, Backer E (1980) Segmentation of muscle cell pictures: a preliminary study. IEEE Trans Pattern Anal Mach Intell PAMI-2(3):232–242
5. Fok YL, Chan JCK, Chin RT (1996) Automated analysis of nerve cell images using active contour models. IEEE Trans Med Imaging 15(3):353–368
6. Ginneken B, Frangi AF, Staal JJ, Haar Romeny BM, Viergever MA (2002) Active shape model segmentation with optimal features. IEEE Trans Med Imaging 21(8):924–933
7. Lie WN (1995) Automatic target segmentation by locally adaptive image thresholding. IEEE Trans Image Process 4(7):1036–1041

8. Couvignou PA, Papanikolopoulos NP, Khosla PK (1993) On the use of snakes for 3-D robotic visual tracking. In: Proceedings of IEEE CVPR, 1993, pp 750–751
9. Kass M, Witkin A, Terzopoulos D (1987) Snakes: active contour models. Int J Comput Vis 1(4):321–331
10. Bahreini L, Fatemizadeh E, Gity M (2010) Gradient vector flow snake segmentation of breast lesions in dynamic contrast-enhanced MR images. In: 17th Iranian conference of biomedical engineering (ICBME2010), 3–4 Nov 2010, pp 1–4
11. Díaz-Parra A, Arana E, Moratal D (2014) Fully automatic spinal canal segmentation for radiation therapy using a gradient vector flow-based method on computed tomography images: a preliminary study. In: 2014 36th annual international conference of the IEEE 26–30 Aug 2014, Engineering in Medicine and Biology Society (EMBC), pp 5518–5521
12. Yuhua Y, Lixiong L, Lejian L, Ming W, Jianping G, Yinghui L (2012) Sigmoid gradient vector flow for medical image segmentation. In: 2012 IEEE 11th international conference on signal processing (ICSP) 21–25 Oct 2012, pp 881–884
13. Qiongfei W, Yong Z, Zhiqiang Z (2015) Infrared image segmentation based on gradient vector flow model. In: 2015 sixth international conference on intelligent systems design and engineering applications (ISDEA), pp 460–462
14. Mahmoud MKA, Al-Jumaily A (2011) Segmentation of skin cancer images based on gradient vector flow (GVF) snake. In: 2011 IEEE international conference on mechatronics and automation, Aug 7–10, Beijing, China, pp 216–220
15. Cleary K (2007) Original datasets. Midas-original datasets. doi:hdl.handle.net/1926/587
16. Al-Kadi OS, Chung DYF, Carlisle RC, Coussios CC, Noble JA (2014) Quantification of ultrasonic texture intra-heterogeneity via volumetric stochastic modeling for tissue characterization. Med Image Anal. https://doi.org/10.1016/j.media.2014.12.004
17. Grove O, Berglund AE, Schabath MB, Aerts HJ, Dekker A, Wang H, Gillies RJ (2015) Data from: quantitative computed tomographic descriptors associate tumor shape complexity and intratumor heterogeneity with prognosis in lung adenocarcinoma. Cancer Imaging Arch. https://doi.org/10.7937/k9/tcia.2015.a6v7jiwx
18. Gonzalez RC, Woods RE, Eddins SL (2009) Digital image processing, using MATLAB, 2nd edn. Gatesmark Publishing, USA
19. Pratt WK (2007) Digital image processing, 4th edn, p 16
20. Pratt WK (1972) Generalized wiener filtering computation techniques. IEEE Trans Comput c-21:636–641
21. Kass M, Witkin A, Terzopolous D (1987) Snakes: active contour models. Int J Comput 1(4):321–331
22. Kazerooni AF, Ahmadian A, Serej ND, Rad HS, Saberi H, Yousefi H, Farnia P (2011) Segmentation of brain tumors in MRI images using multi-scale gradient vector flow. In: 33rd annual international conference of the IEEE EMBS Boston, Massachusetts, USA, August 30–September 3, 2011, pp 7973–7976
23. Xu C, Prince JL (1998) Snakes, shapes, and gradient vector flow. IEEE Trans Image Process 7(3):359

Hyperspectral Image Classification Using Semi-supervised Random Forest

Sunit Kumar Adhikary and Sourish Gunesh Dhekane

Abstract In this paper, a hyperspectral image classification technique is proposed using semi-supervised random forest (SSRF). Robust node splitting in the random forest requires enormous training data, which is scarce in remote sensing applications. In order to overcome this drawback, we propose utilizing unlabeled data in conjunction with labeled data to assist the splitting process. Moreover, in order to tackle the curse of dimensionality associated with a hyperspectral image, we explore nonnegative matrix factorization (NMF) to remove redundant information. Experimental results confirm the efficacy of the proposed method.

1 Introduction

Recent developments in optics and imaging highlight the fact that materials when exposed to various wavelength intervals of electromagnetic spectrum show signature properties [1]. Trees show higher reflectance for wavelengths near infrared region compared with certain visible wavelengths which are its signature property. This behavior is in contrast to other materials like bricks, asphalt, and soil, which have a similar magnitude of reflectance for both ranges of wavelength. A hyperspectral image is basically a collection of images collected in several wavelengths, which often belong to a narrow bandwidth. Thus, it contains a tremendous amount of information and thus finds utility in several remote sensing applications like land cover mapping, change detection, mineral identification, crop analysis, and lot more. Classification of hyperspectral images is the key concept in these applications. However, their low spatial resolution leads to lack of photo-interpretability. This, in turn, hampers the possibility of using supervised learning methods which require labeled data. The acute scarcity of labeled patterns where the number of training samples is rela-

S. K. Adhikary (✉) · S. G. Dhekane
Indian Institute of Information Technology Guwahati, Guwahati, India
e-mail: sunitadhikary1409@gmail.com

S. G. Dhekane
e-mail: sourishdhekane@gmail.com

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_102

tively small compared to the number of features [2] and the curse of dimensionality (i.e., the Hughes phenomenon [3]) pose a serious challenge in the classification of hyperspectral images.

In order to deal efficiently with the problem of lack of labeled patterns, semi-supervised learning (SSL), and active learning (AL) is mostly explored in the remote sensing literature. Semi-supervised techniques try to exploit unlabeled data and extract valuable information from both labeled and unlabeled data during training. On the other hand, active learning techniques build a small set of extremely valuable labeled patterns. A number of SSL techniques have been applied to multispectral and hyperspectral image classification, including Transductive SVMs (TSVM) [4], Laplacian SVMs (LapSVM) [5], and Graph-based methods [6]. Ratle et al. [7] proposed semi-supervised neural networks (SSNN) to label scenes from hyperspectral data which were trained with gradient descent and were shown to outperform traditional margin-based SSL techniques. It has been pointed out in [8] that the effectiveness of margin maximization based methods for SSL depends heavily on specific data distribution which is usually difficult to be satisfied in many applications. Hence, it is desirable to find a method that can utilize the unlabeled data without losing its flexibility.

We explore random forest (RF) for this work. Node splitting in RF is the key issue. Our approach is to use the semi-supervised node splitting criterion stated in [11] to build decision trees. However, the density estimation based technique proposed in [11] seems not to be a good criterion for splitting nodes when there is an acute scarcity of labeled patterns. We have modified it with the incorporation of unsupervised information gain for node splitting. Moreover, nonnegative matrix factorization (NMF) has been used for band reduction and an additional constraint has been imposed in order to be in accordance with the linear mixture model.

2 Methodology

Hyperspectral images are rich in information attributed to its large number of spectral bands. But due to lack of photo-interpretability of hyperspectral images, it is difficult to collect labeled data which poses a serious challenge to the classification of hyperspectral images. Moreover, the spectral bands are in a narrow bandwidth, which leads to a high correlation between several spectral bands. Thus, it becomes absolutely necessary to transform the image into one of the reduced spectral bands. This reduction of spectral bands also considered as feature extraction can be carried out by selecting certain bands or by using a transform that produces features as combinations of bands. This reduction in hyperspectral image analysis can be linked to the hyperspectral unmixing problem. Usually, a pixel in a hyperspectral image contains the reflectance values of various materials because of its low spatial resolution. The observed values at any point can be considered to be a linear mixture of reflectance of materials corresponding to the pixel.

According to the linear mixture model (LMM), each l-dimensional observed pixel vector X present in the image cube can be expressed as a linear combination of "m" materials present in the pixel weighted by its fractional abundance given as

$$X = \sum_{i=1}^{m} u_i v_i + w = Vu + w \quad (1)$$

where V is the $1 \times m$ matrix of spectra (v_1, …, v_m) of the individual composing materials (also called endmembers), u is an m-dimensional vector describing the fractional abundances of the endmembers in the mixture (abundance vector) and w is the additive noise vector. The elements of the abundance vector are assumed to be positive and with unit sum:

$$u_i \geq 0 \; i = 1, \ldots, m \quad (2)$$

$$\sum_{i=1}^{m} u_i = 1 \quad (3)$$

The unmixing can, however, not directly be linked to feature extraction. In techniques like PCA, ICA the endmembers correspond to rows in the linear transform and are orthogonal. This may be a condition too strict for endmember extraction where the base materials may only slightly differ from each other. Moreover satisfying the additivity conditions as stated in Eqs. (2) and (3) cannot be ensured. Hence, we explore NMF in this work which is in coherence with the LMM.

2.1 NMF

NMF is a matrix factorization method which decomposes the given nonnegative matrix $X \in R^{L \times N}$ into a basis matrix ($U \in R^{L \times M}$) and a weight matrix ($V \in R^{M \times N}$) having nonnegative elements such that

$$X \approx UV \quad (4)$$

Appropriate values for U and V are found by formulating an optimization problem by minimizing the Euclidean distance between X and UV, keeping both U and V nonnegative

$$\text{minimize} \quad f(U, V) = \frac{1}{2}\|X - UV\|_F^2 \quad (5)$$

$$\text{subject to} \quad U \geq 0,\ V \geq 0 \quad (6)$$

It should be noted that the operator $\|.\|_{.\mathrm{F}}$ represents the Frobenius norm. The solution of the above optimization problem yields the multiplicative iterative update rules as stated in [9]

$$U_{i,j} \leftarrow U_{ij} \frac{\left(XV^{\mathrm{T}}\right)_{ij}}{\left(UVV^{\mathrm{T}}\right)_{ij}} \tag{7}$$

$$V_{i,j} \leftarrow V_{ij} \frac{\left(U^{T}X\right)_{ij}}{\left(U^{T}XV^{T}V\right)_{ij}} \tag{8}$$

While implementing NMF in this case, the parameter M acts as the hyperparameter. It represents the number of endmembers and is empirically determined. This formulation of NMF does not satisfy Eq. (3). In order to enforce this restriction, we normalize the columns of V after the above optimization problem converges.

2.2 Random Forest

The random forest (RF) is a collection of decision trees $F = \{t_1, t_2, \ldots, t_N\}$ grown on a subset of the training set drawn randomly with replacement also called as bagging [10]. The trees are usually grown until there are only pure samples in a node or until there exists a specified number of samples in it. Data in a node of a tree is split and sent to 2 sister nodes based on a partition criterion like information gain, Gini index, and Bayesian classification error. The splitting of the data can, however, be carried out in various manners [13]. In our work, we chose hyperplanes aligned to the axis. However, in order to invoke more randomness in the decision trees a subset of the feature space is considered in each node [13] to find the best node split. The information gain used to measure the quality of node split is defined as the difference of entropies before and after splitting given as

$$I_j = I_j^s = H(R) - \frac{|R_{\mathrm{l}}|}{|R|} H(R_{\mathrm{l}}) - \frac{|R_{\mathrm{r}}|}{|R|} H(R_{\mathrm{r}}) \tag{9}$$

where R is an internal node, R_{l} and R_{r} are its left and right child, respectively. $H(R)$ is the Shannon entropy of R

$$H(R) = -\sum_{k=1}^{K} p_k \log p_k \tag{10}$$

where p_k is the probability distribution of the kth class in node R of a tree, similarly $H(R_{\mathrm{r}})$ and $H(R_{\mathrm{l}})$ should be calculated. Traditionally, we can use the law of total probability to calculate p_k.

$$p_k = \int_R p(k|x)\mathrm{d}p(x) \tag{11}$$

In the fully supervised training phase which contains labeled samples $p\left(k|x_{R_i}\right)$, which represents the probability of the ith sample in node R having class k is either 1 or 0. So p_k can be readily obtained. While testing a test sample x, RF assigns probability estimate for each class is as follows:

$$p(k|x) = \frac{1}{N}\sum_{i=1}^{N} p_i(k|x) \tag{12}$$

where $p_i(k|x)$ is the probability estimation of class k given by the ith tree. The overall decision function of RF is defined as

$$\mathcal{F} = \operatorname{argmax}_{\mathrm{k}\in\mathrm{y}} p(k|x) \tag{13}$$

2.3 Semi-supervised Node Splitting

The problem with the fully supervised splitting is that, although the distribution $p\left(k|x_{R_i}\right)$ is given by the labeled sample, the sparsely labeled data cannot give a good approximation of the marginal distribution which may lead to a worse choice of the separating hyperplane. Unfortunately, the insufficiency of labeled training data usually leads to a sparse distribution and a bad approximation. Motivated by the work done in [11] we explore unlabeled data to do a semi-supervised node split. A new problem arises, the posteriori distribution $\hat{p}\left(k|x_{R_i}\right)$ of unlabeled data x of node R in the ith tree is unknown. It can be estimated as in [11] as the probability density ratio.

$$\hat{p}\left(k|x_{R_i}\right) = \frac{p\left(x_{R_i}|k\right)}{\sum_{j=1}^{K} p(x_{R_i}|j)} \tag{14}$$

For $p\left(x_{R_i}|k\right)$, we apply a kernel-based density estimation with Gaussian kernel [12]

$$K_h(u) = h^{-d}(2\pi)^{-d/2}\exp\left\{-\frac{1}{2h^2}u^{\mathrm{T}}u\right\} \tag{15}$$

where h is the bandwidth to be determined and d is the data dimension. We then have the following estimation for an unlabeled sample:

$$p\left(x_{R_i}|k\right) = \frac{1}{n_k}\sum_{y_{j=k}} K_h\left(x_{R_i} - x_j\right) \tag{16}$$

where n_k is the number of samples that are labeled k. Once $\hat{p}(k|x_{R_i})$ is determined for the unlabeled samples it can be treated like labeled patterns. Equation (9) can be used to split the nodes, p_k being calculated using Eq. (11). But when there is severely less labeled data in a node, the posterior probability calculated based on the density estimate by the above method cannot be relied upon. In that case, we obtain supervised information gain with the help of a few labeled patterns I_j^{s} as in Eq. (9) to which we add an unsupervised information gain for calculating the total information gain

$$I_j = I_j^{\mathrm{s}} + \lambda\, I_j^{\mathrm{u}} \tag{17}$$

Assuming multi-variate Gaussian distributions at the nodes. The unsupervised gain term I_j^{u} can be calculated with both labeled and unlabeled samples as in [13] is given as

$$I_j^{\mathrm{u}} = \log(\|\Lambda(R)\|) - \frac{|R_{\mathrm{l}}|}{|R|}\log(\|\Lambda(R_{\mathrm{l}})\|) - \frac{|R_{\mathrm{r}}|}{|R|}\log(\|\Lambda(R_{\mathrm{r}})\|) \tag{18}$$

where $\|.\|$ represents determinant and Λ represents the $p \times p$ covariance matrix of the data in the node. The trees are grown on chunks of data bagged from the entire set of labeled and unlabeled samples to an extent until there are pure true labeled samples following the new splitting criterion stated above. The collection of decision trees grown following this scheme is what we refer to as semi-supervised random forest (SSRF).

3 Experiments

3.1 Dataset

A real-world hyperspectral image dataset called "University of Pavia" (UP) is used to validate the efficacy of the proposed method in our study. Reflective Optics System Imaging Spectrometer (ROSIS-3) sensor has been used to collect the dataset. It has 103 spectral bands collected within the wavelength range of 430–860 nm. The image consists of 610 × 340 pixels at a spatial resolution of 1.3 m per pixel. This dataset considers 9 land cover classes of interest. Table 1 shows a number of samples in each land cover class.

Table 1 Number of samples per class

Class labels	Number of samples
Asphalt	6631
Meadows	18,649
Gravel	2099
Trees	3064
Painted metal sheets	1345
Bare soil	5029
Bitumen	1330
Self-blocking bricks	3682
Shadows	947

3.2 Experimental Setup and Results

In our experiment, the given hyperspectral image is decomposed using the NMF scheme described above. We perform 300 iterations with the loss function converging to a global/local optima. The number of endmembers has been empirically chosen to be 10. For the classification process, labeled set $\boldsymbol{X}_{\mathbf{L}}$ contains 5, 10, 15 samples per class, respectively, randomly selected from the labeled pixels on the hyperspectral image. The test set includes all the samples leaving aside the samples used for training the classifier. The decision trees of semi-supervised random forest (SSRF) are grown on the entire dataset (both training and test data). Each node in a tree is split with the semi-supervised node split criterion described above. When the number of labeled samples in a node is more than 4, we estimate the posterior probability of unlabeled data with the density estimation step as is stated in Eq. (16) and then use the supervised information gain stated in Eq. (9) to do the node split as if the entire data was a supervised data. But when the number of labeled samples is severely less in a particular node the density estimated with Gaussian kernel cannot be relied upon and hence we add an unsupervised information gain term to the supervised information gain as in Eq. (17). Several values of λ were experimented and was finally set to 0.3. The supervised information gain is calculated with the small amount of labeled patterns in the node but the entire labeled and unlabeled patterns are used to obtain the unsupervised information gain. The generated SSRF consist of a total of 300 trees.

The classification accuracy of the proposed technique is compared with TSSNN [7] and LapSSNN [7]. The results are shown in Table 2. The proposed technique can be seen to beat the neural network based semi-supervised schemes with 10 and 15 samples per class but gives a comparable result in 5 sample per class. Thus, it can be concluded that the proposed scheme is better able to leverage information present in unlabeled data.

Table 2 Classification accuracy for Pavia University dataset

Labeled examples per class	5	10	15
TSSNN [7]	**73.2 ± 2.79**	75.3 ± 2.09	76.3 ± 5.24
LapSSNN [7]	72.7 ± 2.82	74.1 ± 1.35	74.9 ± 4.07
Proposed technique	72.9 ± 3.72	**77.6 ± 1.66**	**79.2 ± 1.1**

4 Conclusion

In this paper, we proposed a semi-supervised random forest for classification. The improved semi-supervised node split criterion has successfully leveraged information from unlabeled data. The enhanced value of classification accuracy and its low variance bear testimony of the effectiveness of the proposed technique. Apart from the improvement in classification accuracy, the proposed technique is computationally efficient. Unlike traditional SSL techniques which are either iterative in nature or are optimized with gradient descent, the proposed semi-supervised technique has less time complexity when expressed as a function of the number of samples.

Acknowledgements The authors thank Prof. Gamba for providing the University of Pavia dataset.

References

1. Mather PM, Koch M (2010) Computer processing of remotely-sensed images, 4th edn. Wiley, New York
2. Baraldi A, Bruzzone L, Blonda P (2005) Quality assessment of classification and cluster maps without ground truth knowledge. IEEE Trans Geosci Remote Sens 43:857–873
3. Hughes GF (1968) On the mean accuracy of statistical pattern recognition. IEEE Trans Inf Theory 14:55–63
4. Bruzzone L, Chi M, Marconcini M (2006) A novel transductive SVM for semisupervised classification of remote-sensing images. IEEE Trans Geosci Remote Sens 44:3363–3373
5. Gomez-Chova L, Camps-Valls G, Munnoz-Mari J, Calpe J (2008) Semisupervised image classification with Laplacian support vector machines. IEEE Geosci Remote Sens Lett 5:336–340
6. Camps-Valls G, Bandos TV, Zhou D (2007) Semi-supervised graph-based hyperspectral image classification. IEEE Geosci Remote Sens Lett 45:3044–3054
7. Ratle F, Camps-Valls G, Weston J (2010) Semisupervised neural networks for efficient hyperspectral image classification. IEEE Trans Geosci Remote Sens 48(5):2271–2282
8. Zhang T, Oles F (2000) A probability analysis on the value of unlabeled data for classification problems. In: 17th International conference on machine learning. ACM Press, California, pp 1191–1198
9. Lee DD, Seung S (2000) Algorithms for non-negative matrix factorization. In: Neural information processing systems. MIT Press, Denver, pp 556–562
10. Breiman L (1996) Bagging predictors. Mach Learn 26:123–140
11. Liu X, Song M, Tao D, Liu Z, Zhang L, Chen C, Bu J (2013) Semi-supervised node splitting for random forest construction. In: Conference on computer vision and pattern recognition. IEEE Press, Oregon, pp 492–499

12. Silverman B (1986) Density estimation for statistics and data analysis. Chapman and Hall/CRC, London
13. Criminisi A, Shotton J, Konukoglu E (2011) Decision forests for classification, regression, density estimation, mainfold learning and semi-supervised learning. Technical report, Microsoft Research

Clustering of Various Parameters to Catalog Human Bone Disorders Through Soft Computing Simulation

S. Ramkumar and R. Malathi

Abstract Every minute nearly 20 fractures occur due to bone disorders in the world. People around the world could not able to differentiate the difference between bone disorders. This chapter is a novel approach toward differentiation of different bone disorders like osteoporosis and osteopenia with influences several parameters. Accordingly, five different parameters such as Calcium, Phosphate, Vitamin D3, Parathyroid hormone (PTH) level, and calcitonin level are considered for the study to categorize the bone disorders. The present approach is an attempt to combine the clinical data measured from each patient and their respective bone mineral density value for the better classification. This is a unique study to provide combined information of both clinical and image processing studies. For this purpose, the above-mentioned parameters and bone density values were observed from ten different patients. All these data were used as an input for soft computing using MATLAB for further processing the data. Initially, unsupervised mapping classifier is adopted to classify bone disorder, for which the clinical parameters are compared with bone density value using k-means clustering algorithm. The prime idea behind using of k-means technique is that the feasibility to classify the inputs based on the distance between the input seeds. With reference to the perpendicular distance between the seeds inputs, the bone disorders have been cataloged. The repeated iterations lead to best clustering results.

S. Ramkumar (✉) · R. Malathi
Department of EIE, Annamalai University, Chidambaram, Tamil Nadu, India
e-mail: sram0829@gmail.com

R. Malathi
e-mail: vsmalu@gmail.com

S. Ramkumar
Department of EIE, Veltech University, Chennai, Tamil Nadu, India

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_103

1 Introduction

Bone is an essential material of the human skeleton to enable locomotion and ensuring the protection of living organisms. The main function of skeleton bone is to provide mechanical support and also in maintaining calcium homeostasis and hematopoiesis of an organism. When compared to men, women are three times prone to osteoporosis that leads to a femur bone mass deficiency after menopause. The role of hormonal change after menopause in women is inevitable in inducing osteoporotic conditions [1, 2]. Femur bone is the longest and strongest bone in our body, extended from hip to the knee. Particularly, femur that is thigh bone will be affected more due to frequent physical activities such as walking, running, standing, etc.

The human body is made up of different kinds of tissues such as muscle tissue, nervous tissue, epithelial tissue, and connective tissue. The connective tissue is a type of tissue which is responsible for the formation of bones. The subdivisions of the connective tissue are proper connective tissue and specialized connective tissue. Particularly, the bone tissues belong to specialized connective tissue category and appeared in vascularization state with a hard consistency. Due to continuous mineral deposition (mainly calcium) and collagen fibers in extracellular matrix raised the hardness of bone. In general, the human body consists of 206 bones which are seen in different shapes and sizes. Around 80% of bone mass in the skeleton is composed of cortical bone. At the beginning of nineteenth century, the term osteoporosis was proposed to describe the porosity of an elderly human based on histological findings.

Osteoporosis is a quite common bone disease in elderly people of both genders mainly due to comprised bone strength. This has lead to an increased risk of fracture. In women, postmenopausal osteoporosis is highly responsible for such decrease in bone density and bone strength. BMD (bone mineral density) can be measured by following methods such as SPA (Single Photon Absorptiometry) QCT (Quantitative Computed Tomography), QUS (Qualitative Ultrasound), (DPA) Dual Photon Absorptiometry, DXR (Digital X-ray Radiogrammetry), SEXA (Single-energy X-ray Absorptiometry), and DEXA (Dual-energy X-ray Absorptiometry). These tests are carried out for various diseased conditions like Hypophosphatasia, Klippel Feil syndrome, Craniosynostosis, Fibrous dysplasia, and Osteoarthritis involved with bone mineral density reduction [3–6].

In the present study, DEXA method output has been selected to conduct a comparative assessment of associated clinical parameters [7] output for bone disease patients'. Based on the comparative assessment, the classification of the disease occurs into their sub levels. This study will give additional information on patient classification and fracture-prone area to prevent or beneficial treatment of patients. The most successful practice for measuring bone mass density (BMD) is dual-energy X-ray absorptiometry (DEXA) scanning [8, 9]. The output of the DEXA is represented as T-score system as derived from comparing the individual BMD measurement of patients with normal. Osteoporosis is predominantly found in women rather than men, minimum prone to bone fracture because of their high bone density. Based on the literature survey, osteoporosis is less or no symptomatic disease. It is quite difficult to identify this disease at an early stage. After clinical manifestation and occurrence of noticeable loss around 30–40% in bone mass of patients are considered as a diagnostic parameter. Due to limitations in early diagnostic procedures, there is no prevention of this disease at this stage. All the supportive treatment of osteoporosis patients can support in reducing the severity of bone density loss instead of curing the disease. Therefore, this disease is termed as "silent epidemic" disease. Especially, age and gender are non-modifiable risk factors responsible for both osteopenia and osteoporosis condition. Unless changes in bone density, this disease is unpredictable [10, 11].

Osteoporosis has been considered as a public health problem, its diagnosis and fractures rehabilitation are high and it shed an impact on treatment cost. It is such a burden for developing countries to meet their countries growth and health status. To avoid severe bone loss and fractures, early diagnosis method is the only solution for these types of bone diseases. This study focused to develop combine diagnostic methods using a computational algorithm and its application for identification and classification of bone disease patients into either osteopenia or osteoporosis condition using input values of clinical diagnosis and iDEXA scanning method. In the case of medical application, sensitivity in diagnostic procedures and medical problem detection and its accuracy all must be taken into consideration.

There are many types of fluids available in the human. In all those body fluids, calcium is the key electrolytes in body fluids. In general, 99% of calcium intake is from the food and it is utilized for bone mineralization. The remaining 1% is inevitable for neuromuscular excitability, cardiac activity, blood coagulation, and membrane permeability [12]. The calcium in the body (e, i) present in plasma either surrounded by albumin or it may be presented as in free ionized form. Serum calcium level normal ranges from 8.8 to 10.4 mg/dl in a healthy adult. The total free calcium present in three forms such as calcium ion account, protein-bound complex, and ionic complex. At the same time, the percentage of calcium contents is also varied as 51, 40, and 9%. Likewise normal physiological phosphate ranges from 0.8 to 3.74 mg/dl for healthy adult; [13] The vitamin D3 level ranges from 30 to 100 ng/mL; Parathyroid hormone (PTH) level ranges from 1.9 to 106 pg/mL; [14] and the normal range in calcitonin level is different for male and female that is for female (0.1–10.9) and for male (0.2–27.7) [15].

2 Research Methodology

The research methodology completely depends on the classification of bone disorders. The inputs are classified based on the training and further, it helps to cluster them into groups. In this classification, clustering is done using k-means clustering algorithm.

Table 1 Serum calcium, Ph, Vit-D3, PTH, Calcitonin, and DEXA level inputs [16]

S. No.	Sex	Age	Measured serum level (Ca) (mg/dL)	Measured serum level (Ph) (mg/dL)	Measured serum level (Vit-D3) (ng/mL)	Measured serum level (PTH) (pg/mL)	Measured serum level (Calci-tonin)	DEXA value
1	F	55	8.31	5.9	20.3	13	2.99	−1.893
2	F	43	7.90	4.07	21.4	28	3.11	−1.661
3	M	69	8.25	8.31	29.9	37	10.14	−1.931
4	M	58	8.19	6.60	28.5	33	9.33	−1.895
5	F	55	5.94	10.47	20.5	11	0.07	−2.574
6	F	52	6.33	9.55	15.8	19	2.29	−1.417
7	M	78	7.31	8.11	25.1	22	0.20	−1.768
8	F	51	8.22	9.39	27.0	14	0.88	−1.915
9	F	66	7.93	8.52	22.8	17	2.03	−1.355
10	F	72	5.65	7.99	19.9	12	0.05	−2.901

2.1 k-Means Clustering Algorithm

k-means clustering is a machine learning algorithm or data mining, which is used to cluster the data into clusters of related data without any information of those associations. It is one of the clustering methods and it is normally used in biometrics, medical imaging, etc. It is an evolutionary algorithm that the name set from its method of operation. The algorithm clusters data into k groups, where k is an input parameter. Then allocate every observation to clusters based on the observation's closeness to the mean of the cluster. The clusters mean is then re-estimated and the procedure initiates again. The k-means algorithm works as follows:

1. The algorithm randomly picks k points as the primary cluster centroids.
2. Every point in the dataset is allocated to the closed cluster based on the Euclidean distance between every point and cluster center.
3. Each cluster centroid is re-estimated as the average of the points in each cluster.
4. Steps 2 and 3 are repeated waiting for the clusters to diverge. Convergence is well defined depending upon the implementation, but it usually means that either no observations alteration clusters when steps 2 and 3 are repeated or that the variations do not make a material modification in the definition of the clusters [16] (Table 1).

3 Data Collection

The k-means clustering is done in order to find the nearest clusters for Calcium, Phosphate, Vitamin D3, PTH, and Calcitonin. In this, the number of datasets that is being adapted is 10, i.e., Calcium has 10 values which include the age, gender, serum

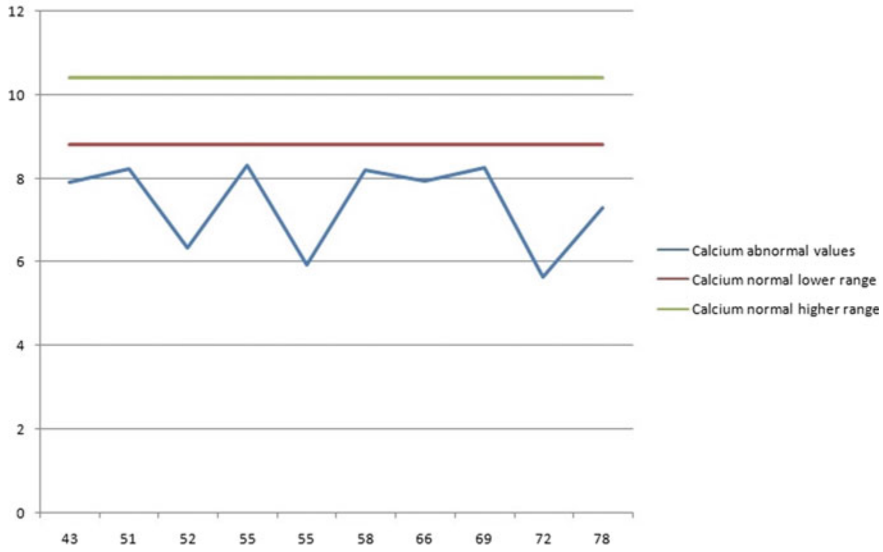


Fig. 1 Graph between abnormality and normal for Calcium

level, and T-score value similarly the values has been detected for Phosphate, Vitamin D3, PTH, and Calcitonin. The following figures were plotted for representation of the abnormalities of Calcium, Phosphate, Vitamin D3, PTH, and Calcitonin. The following plot representation the threshold values of parameters Figs. 1, 2, 3, 4, and 5 [17, 18].

4 Result with Discussion

The function k-means partitions data into k equally exclusive groups and proceeds the index of the cluster to which it has allotted every reflection [16]. Unlike hierarchical clustering, k-means clustering functions on actual observations and produces a single level of clusters. The partitions mean that k-means grouping is often extra fit than hierarchical clustering for huge amounts of data. k-means works for every observation in given data as an object having a location in space. It determines a partition in which substances within each cluster are as close to each other as possible and as far from substances in other clusters as possible. Every cluster in the partition is defined by its member objects and by its centroid or center [19]. The centroid for every cluster is the point to which the sum of distances from all substances in that cluster is reduced. This has been obtained in the Figs. 6, 7, 8, 9, and 10.

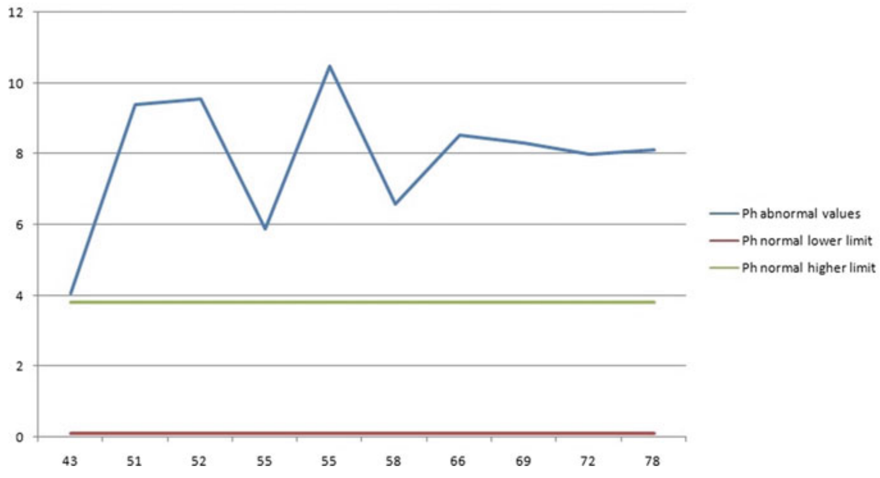


Fig. 2 Graph between abnormality and normal for Ph

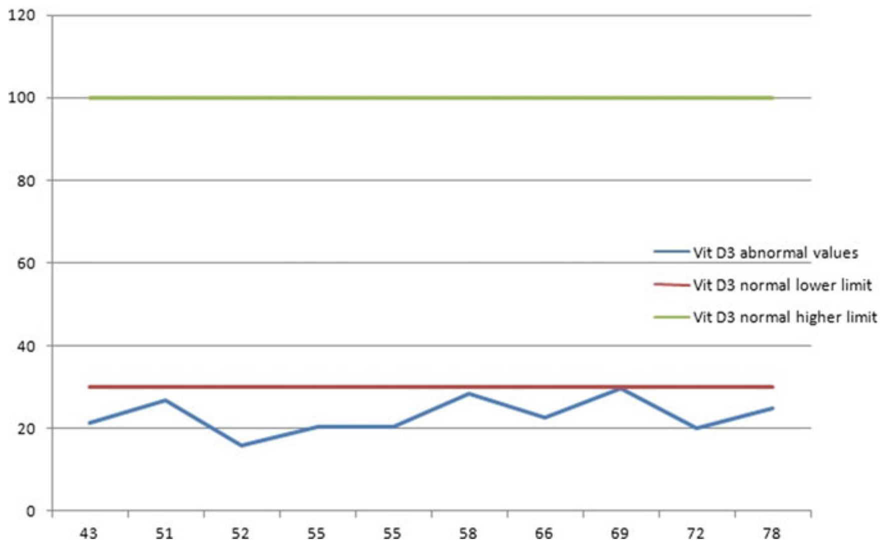


Fig. 3 Graph between abnormality and normal for Vitamin D3

4.1 Calcium

See Fig. 6.

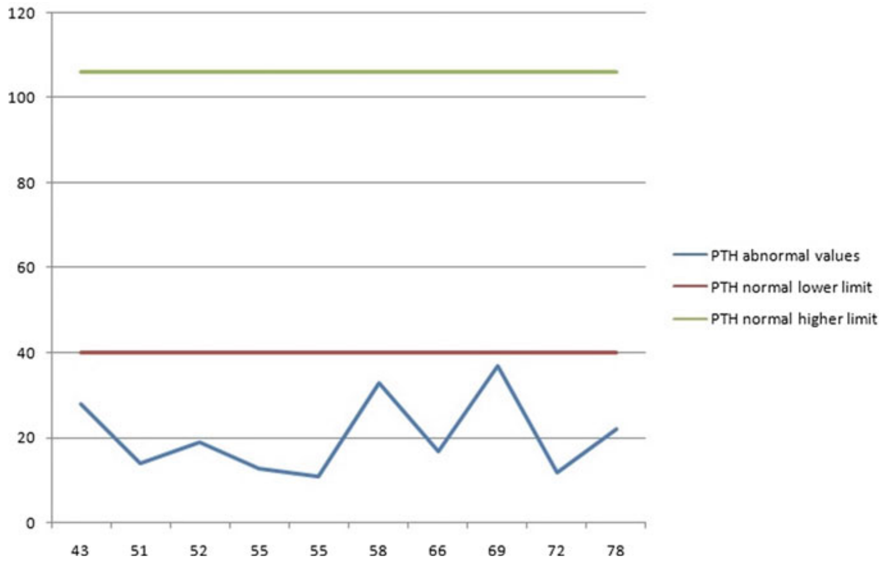


Fig. 4 Graph between abnormality and normal for PTH

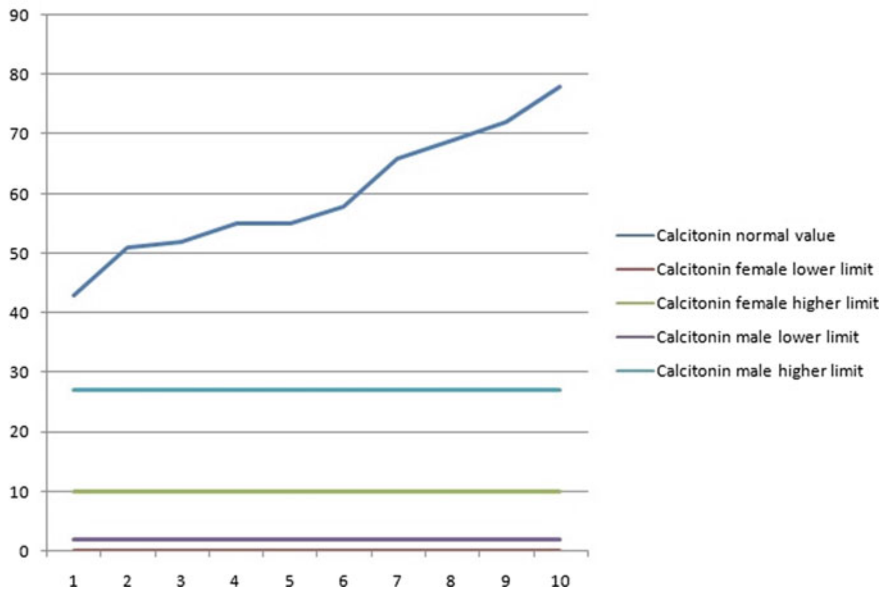


Fig. 5 Graph between abnormality and normal for Calcitonin

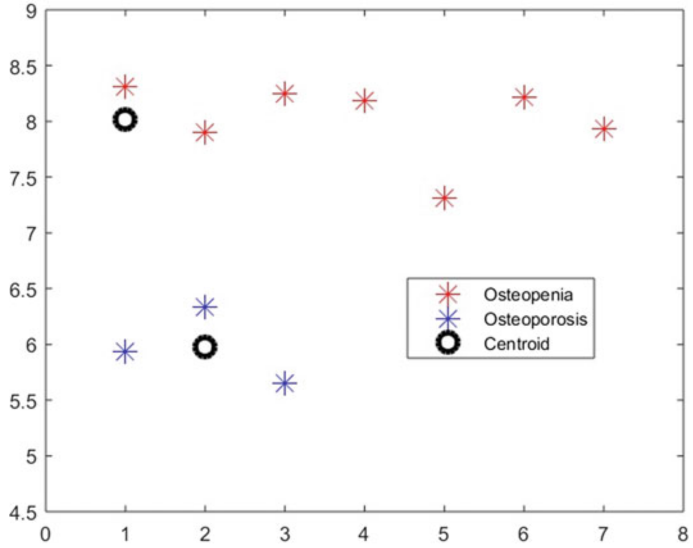


Fig. 6 Output of k-means clustering of Ca

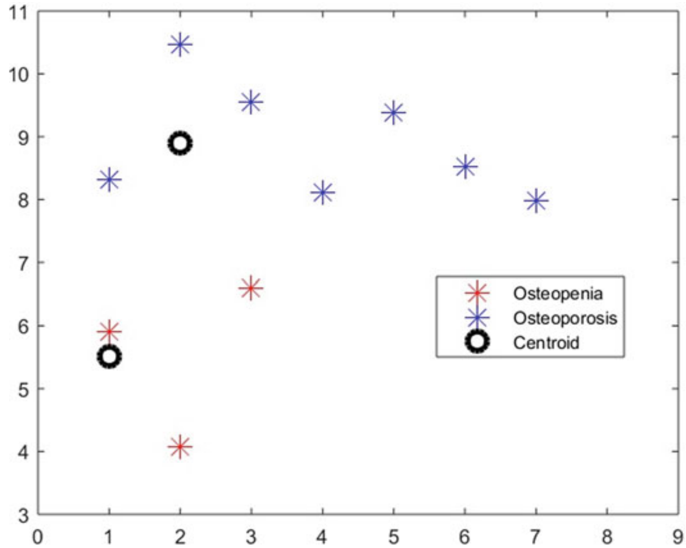


Fig. 7 Output of k-means clustering of Ph

4.2 Phosphate

See Fig. 7.

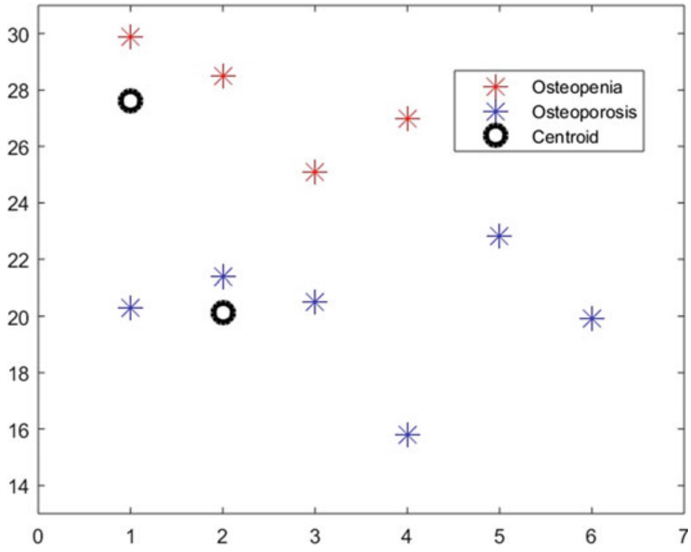


Fig. 8 Output of k-means clustering of Vitamin D3

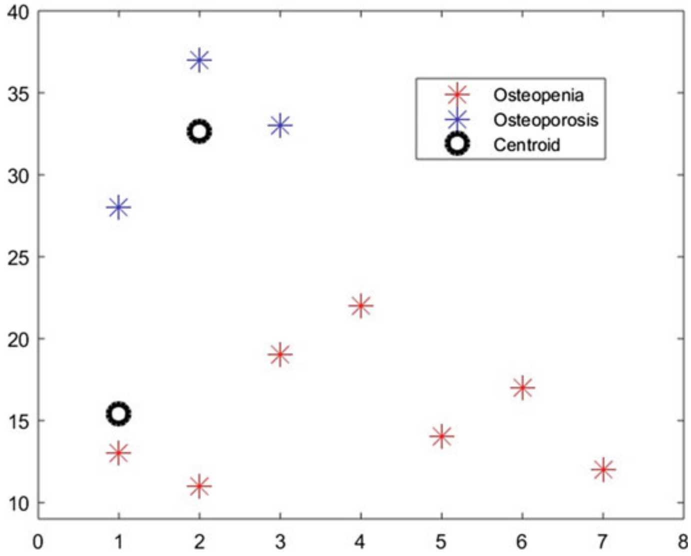


Fig. 9 Output of k-means clustering of PTH

4.3 Vitamin D3

See Fig. 8.

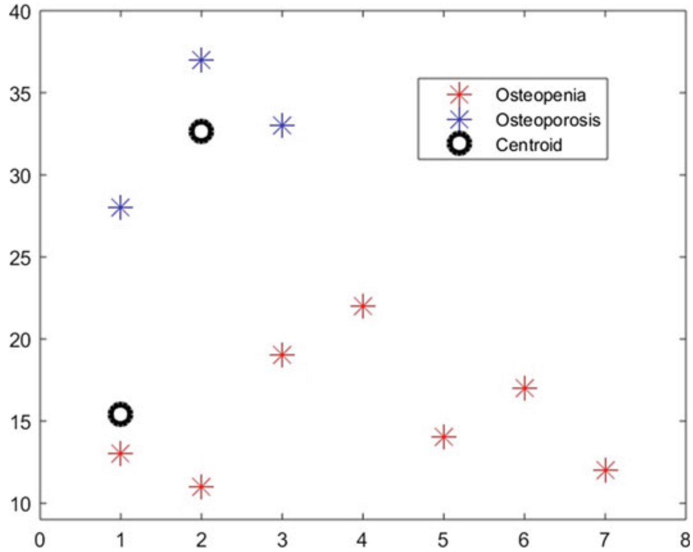


Fig. 10 Output of k-means clustering of PTH

4.4 PTH

See Fig. 9.

4.5 Calcitonin

The number of inputs in a two-dimensional plane will be clustered into different groups based on 'k' value. Select any two inputs from entire inputs. The selected inputs were named as seed inputs. A geometrical line is drawn between the seed inputs, which is bisected by a perpendicular line. The bisected line divides the entire inputs into two half, i.e., both sides of the perpendicular line form two different groups. To obtain optimal results, iterations were carried out. Hereby, the 'k' value assigned as 2, so the two groups of clusters are formed based on the bone diseases. Hence, the cluster's inputs were noted with different colors such as blue and red.

5 Conclusion

The present approach has attempted to combine clinical data measured from each patients and their respective DEXA value for classification of patients into different catalogues of bone disorder conditions. This approach has included 10 patients and their five influencing parameters such Calcium, Phosphate, Vitamin D_3, Parathyroid hormone (PTH) level and calcitonin level are taken into account along with respective DEXA value. The k-means clustering algorithm is a type of neural network. Particularly, this algorithm uses multilevel networks. The number of inputs in a two-dimensional plane will be clustered into different groups based on 'k' value. Select any two inputs from entire inputs. The selected inputs were named as seed inputs. A geometrical line is drawn between the seed inputs, which is bisected by a perpendicular line. The bisected line divides the entire inputs into two half, i.e., both sides of the perpendicular line form two different groups. To obtain optimal results, iterations were carried out. Hereby the 'k' value assigned as 2, so the two groups of clusters formed based on the bone diseases. Hence the cluster's inputs were noted with different colors such as blue and red. This results in the classification of bone disorders based on severities. This concludes that each one of the parameters creates an impact on the results. In case of any abnormalities at one of the parameters' values then it can be considered as the beginning stages of bone disorder diseases. Further testing this soft simulation with an increased number of population, this method could be improvised. Therefore, the application of this method can be extended. This approach creates awareness among the people toward bone disorder diseases.

References

1. Shatrugna V, Kulkarni B, Kumar PA, Rani KU, Balakrishna N (2005) Bone status of Indian women from a low-income group and its relationship to the nutritional status. Osteoporos Int 16:1827–1835. https://doi.org/10.1007/s00198-005-1933-1
2. Siris ES, Miller PD, Barrett-Connor E, Faulkner KG, Wehren LE, Abbott TA et al (2001) Identification and fracture outcomes of undiagnosed low bone mineral density in postmenopausal women. JAMA 286:2815. https://doi.org/10.1001/jama.286.22.2815
3. Park M, Kang B, Jin SJ, Luo S (2009) Computer aided diagnosis system of medical images using incremental learning method. Expert Syst Appl 36:7242–7251. https://doi.org/10.1016/j.eswa.2008.09.058
4. Ding F, Leow WK, Sen Howe T (n.d.) Automatic segmentation of femur bones in anterior-posterior pelvis X-ray images. In: Computer vision analysis of images patterns. Springer, Berlin, Heidelberg, pp 205–212. https://doi.org/10.1007/978-3-540-74272-2_26
5. Lim SE, Xing Y, Chen Y, Leow WK, Sen Howe T, Png MA (2004) Detection of femur and radius fractures in X-ray images. In: Proceedings of the 2nd international conference on advanced medical signal information processing, vol 1. pp 249–256
6. Armato III SG, Sensakovic WF (n.d.) Automated lung segmentation for thoracic CT: impact on computer-aided diagnosis 1. https://doi.org/10.1016/j.xacra.2004.06.005
7. Pulkkinen P, Jämsä T, Lochmüller E-M, Kuhn V, Nieminen MT, Eckstein F (2008) Experimental hip fracture load can be predicted from plain radiography by combined analysis of trabecular bone structure and bone geometry. Osteoporos Int 19:547–558. https://doi.org/10.1007/s00198-007-0479-9
8. Sapthagirivasan V, Anburajan M (2013) Diagnosis of osteoporosis by extraction of trabecular features from hip radiographs using support vector machine: an investigation panorama with DXA. Comput Biol Med 43:1910–1919. https://doi.org/10.1016/j.compbiomed.2013.09.002

9. Kavitha MS, Asano A, Taguchi A, Kurita T, Sanada M (2012) Diagnosis of osteoporosis from dental panoramic radiographs using the support vector machine method in a computer-aided system. BMC Med Imaging 12:1. https://doi.org/10.1186/1471-2342-12-1
10. Detection of osteoporosis and osteopenia using bone densitometer – simulation study. Materials Today: Proceedings (Elsevier) Volume 5, 1024–6
11. Marwaha RK, Tandon N, Garg MK, Kanwar R, Narang A, Sastry A et al (2011) Bone health in healthy Indian population aged 50 years and above. Osteoporos Int 22:2829–2836. https://doi.org/10.1007/s00198-010-1507-8
12. Wang L, Nancollas GH, Henneman ZJ, Klein E, Weiner S (2006) Nanosized particles in bone and dissolution insensitivity of bone mineral. Biointerphases 1:106–111. https://doi.org/10.1116/1.2354575
13. World Health Organization (2004) WHO scientific group on the assessment of osteoporosis at primary health care level. In: Summary meeting report, pp 5–7
14. Nalavade K, Meshram BB (2014) Evaluation of k-means clustering for effective intrusion detection and prevention in massive network traffic data. Int J Comput Appl 96(7):9–14
15. Kavitha MS, Asano A, Taguchi A, Kurita T, Sanada M (2012) Diagnosis of osteoporosis from dental panoramic radiographs using the support vector machine method in a computer-aided system. BMC Med Imaging 12(1):1
16. K*-means: an effective and efficient k-means clustering algorithm. IEEE Xplore. Retrieved from https://doi.org/10.1109/bdcloud-SocialComSustainCom.2016.46. Accessed on 31 Oct 2016
17. McCormick CC (2002) Passive diffusion does not play a major role in the absorption of dietary calcium in normal adults. J Nutr 132:3428–30. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12421863
18. Kanis JA (2004) WHO scientific group on the assessment of osteoporosis at primary health care level. World Health Organisation, 5–7 May 2004. https://doi.org/10.1016/s0140-6736(02)08761-5
19. Vesanto J, Alhoniemi E (2000) Clustering of the self-organizing map. IEEE Trans Neural Netw 11:586–600. https://doi.org/10.1109/72.846731

IoT-Based Embedded Smart Lock Control Using Face Recognition System

J. Krishna Chaithanya, G. A. E. Satish Kumar and T. Ramasri

Abstract Smart home security and remote monitoring have become vital and indispensable in recent times, and with the advent of new concepts like Internet of Things and development of advanced authentication and security technologies, the need for smarter security systems has only been growing. The design and development of an intelligent web-based door lock control system using face recognition technology, for authentication, remote monitoring of visitors and remote control of smart door lock have been reported in this paper. This system uses Haar-like features for face detection and Local Binary Pattern Histogram (LBPH) for face recognition. The system also includes a web-based remote monitoring, an authentication module, and a bare-bones embedded IoT server, which transmits the live pictures of the visitors via email along with an SMS notification, and the owner can then remotely control the lock by responding to the email with predefined security codes to unlock the door. This system finds wide applications in smart homes where the physical presence of the owner at all times is not possible, and where a remote authentication and control is desired. The system has been implemented and tested using the Raspberry Pi 2 board, Python along with OpenCV are used to program the various face recognition and control modules.

1 Introduction

With the advent of various smart technologies, the need for better and more intelligent security and monitoring has been growing. In recent times, the need for security and surveillance has become vital in many areas such as homes, offices, banks, etc. In the recent past, various authentication techniques have been designed and implemented, passwords, patterns, RFID to name a few, these technologies have

J. Krishna Chaithanya (✉) · G. A. E. Satish Kumar
Department of ECE, Vardhaman College of Engineering, Hyderabad, India
e-mail: j.krishnachaitanya@vardhman.org

T. Ramasri
Department of ECE, SVUCE, SVU, Tirupati, India

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_104

their advantages and disadvantages, the passwords and patterns once traced may compromise the security, the fingerprint-based systems are regularly prone errors arising due to external factors and mismatches. Hence, arises a need for a more efficient and effective way of authentication [1]. The authentication based on Face Recognition has played a pivotal role over the years providing unmatched levels of efficiency and accuracy [2, 3] but was limited to a few high-security establishments and large corporations as the design, and implementation costs are high. Today, thanks to ongoing research and development, the algorithms have become more accessible and find a broad range of applications [4]. Also, the present day home security systems have not been updated since years, the physical presence of the owner/key is mandatory to gain access to the house, the proposed system hence includes an intelligent web-based embedded server aimed at providing remote authentication and control of the door lock using email and basic IoT concepts.

1.1 Face Recognition

In the proposed system, we introduce a low-cost extendable framework for embedded smart home security system, which consists of a face recognition module. This system uses Haar-like features for face detection and Local Binary Pattern Histogram (LBPH) for face recognition [5, 6]. The system a cascade classifier in face detection and face recognition is carried out in three stages, namely feature extraction, matching, and classification. The distinctive and at the most important options are extracted, and the face image is compared with pictures in information throughout the last stage (classification). This native binary pattern for person's face recognition takes into account each form and texture information for analysis. The image given is segregated into little elements from which the Local Binary Patterns are adopted and clubbed into one vector feature. This feature vector helps in measuring similarities between pictures by forming an associate in nursing economical illustration of a face. The algorithms are implemented using the OpenCV library of Python programming language running on the Raspberry Pi 2.

1.2 Remote Monitoring Control

In addition to the face recognition-based authentication, the proposed system also includes an intelligent web-based embedded server for remote authentication and control of the door lock. This technology plays a crucial role in enabling remote access to a person/persons if required and if deemed necessary by the owner, who is not required to be present in-person. When an unauthorized person is detected by the system, it emails the owner a live picture of the individual along with as SMS notification and waits for his command. If the owner recognizes the person and would like to provide them with access, he can do so by sending a predefined security code to the server via email, the server, in turn, checks the authenticity of the code and unlocks the door accordingly [7, 8]. The server also maintains a web page which can only be accessed by the owner to bypass the face recognition system in case of error. This subsystem is implemented using Python programming language along with the Linux server and PHP scripts running on the Raspberry Pi 2.

2 System Design and Architecture

2.1 System Architecture

The proposed system is a combination of various modules namely, imaging module, core module and the door lock module. Figure 1 represent the block diagram. The imaging module is responsible for capturing the images of visitor/visitors and forwarding it to the core module for further processing; this is realized using a USB web camera. The door lock module is a combination of driver circuitry and a DC motor to drive a lock, responsible for locking and unlocking the door as required [4] (Fig. 2).

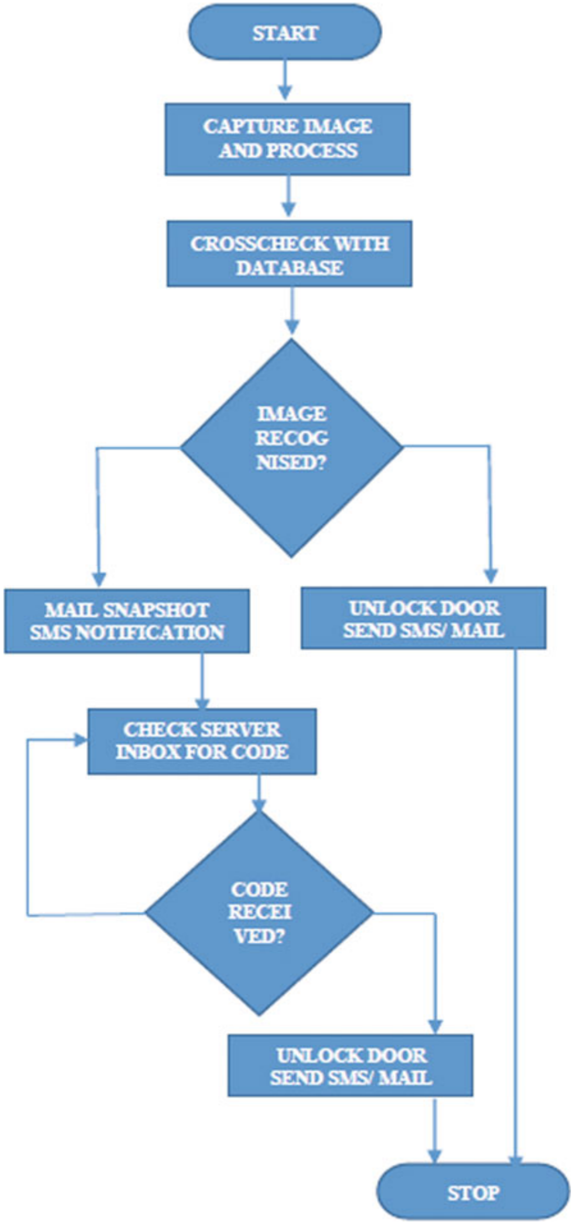
The heart of this system is the core module which is realized using the Raspberry Pi 2, its responsibilities include, acquiring images from the camera, processing the acquired image as required, maintaining the facial image database, comparing the acquired image with the database, sending commands to the door lock module, etc. The core module also acts as an embedded web server responsible for sending and receiving emails, sending out SMS notifications and backend access [7].

3 System Description

3.1 Imaging Module

The imaging module in the proposed system is realized using a USB web Camera, the main reason behind choosing USB Camera over the Pi camera is the cost effectiveness. The camera features a high-quality CMOS sensor, with an image resolution of 25 MP (Interpolated), an adjustable lens for focus adjustment, a frame rate of 30 fps and f2.0 lens. The USB camera also is equipped with night vision for low light photography. The camera interfaces with the Raspberry Pi via the USB 2.0 port and is responsible for capturing images when requested, the pictures are captured by using the command fswebcam (Fig. 3).

Fig. 1 Flow chart



3.2 *Raspberry Pi Core Module*

The core module of the system is realized using a Raspberry Pi 2 board; it's a $35 bare-bones computer designed and developed by the Raspberry Pi Foundation, the Pi

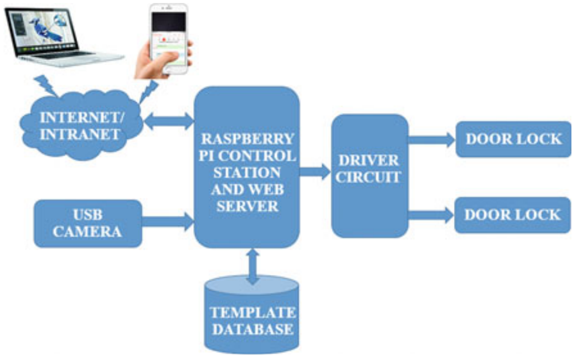
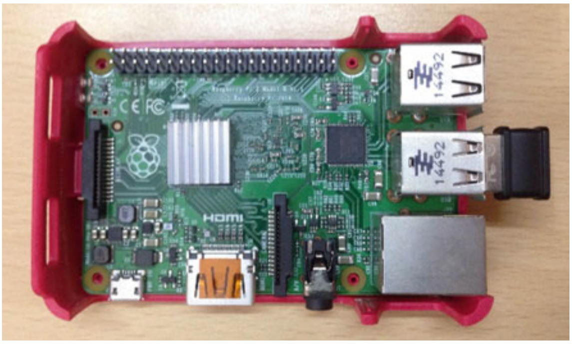


Fig. 2 Block diagram

Fig. 3 USB camera

2 features a BCM 2836 System-on-Chip which includes a Quad-Core 32-Bit ARM Cortex A7 CPU clocked at 900 MHz paired with 1 GB of RAM. It also has VideoCore IV GPU for graphical processing applications, it also includes four USB ports for peripherals and 40 Pin General Purpose Input Output (GPIO) pins for interfacing the Pi with external electronic circuits, these GPIO pins are used to interface the Pi to the door lock module. The Raspberry Pi is designed to run various Linux based operating systems and has Raspbian as its official operating system and Python as its official programming language. In this system the core module plays a highly pivotal role and is responsible for various functions, the core module is responsible for acquiring the images from the camera, processing and storing. It's also responsible for maintaining the facial database which consists of pictures of all the authorized persons for reference. It is in charge of employing the face detection and recognition algorithms and has to decide whether is a person is authorized or not. It's responsible for controlling the door lock module by sending lock/unlock commands using Python code via GPIO to the motor driver (Fig. 4).

Fig. 4 Raspberry Pi 2 module



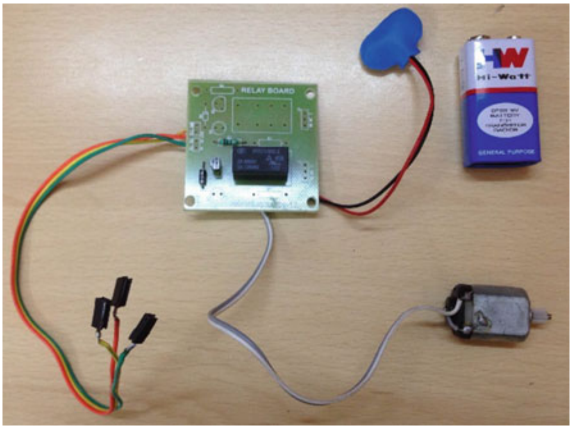
3.3 Embedded Server and IoT

Another crucial function of the core module is to act as an embedded web server, the primary responsibilities of this server include, transmitting the visitor/visitors images via email to the owner, SMS notifications, look for emails from the owner and find the security code from the emails for authorization. This system employs an embedded server approach for communicating with the user and with the internet/intranet. Python code is used to program certain aspects of this system such as sending and receiving emails and text messages. Standard Python libraries corresponding to the web such as urlib2, cookielib for online SMS service; imaplib, poplib, email, smtp, etc. for sending and receiving emails are imported and used accordingly. This system uses web-based SMS client way2sms to send SMS alerts to the owner; it's also configured with a dedicated gmail id to send and receive emails. The system is also configured using Apache to act as a server, which is useful to remotely monitor the conditions. The owner can log into the server using a dedicated static IP assigned to the Raspberry Pi, another important function of this server is to provide a secure back door to lock/unlock the door by bypassing the face recognition feature in case of a failure or emergency. This is a secret feature and is only accessible by the owner.

3.4 Face Detection and Recognition

Many kinds of face detection algorithms are used in many appliances, surveillance systems, gaming, human-computer interaction, etc. Paul Viola and Michael Jones devised a formula for object detection using Haar-feature based cascade classifiers. It's a machine learning based algorithm in which some positive and negative images are employed to train the cascade classifier. Once the classifier is trained, features are extracted which is in turn used for object/face detection [5]. The OpenCV library of Python provides support for using the Haar cascade classifiers for face detection, and it's equipped with both trainer and classifier. Thus the required XML classifiers

Fig. 5 Door lock module



are loaded using Python. For face recognition to be carried out, various face recognition algorithms such as Eigenfaces, Fisherfaces, LBPH Algorithm, etc. are available [6, 10]. This system uses Lower Binary Pattern Histograms method to perform face recognition. The Eigenfaces and Fisherfaces methods employ a holistic approach to recognitions, the data in these techniques are treated as a vector in high dimensional space which is not always ideal, hence in the LBPH algorithm the idea is to look at the lower dimensional subspace for useful information. In this method, the local structure in an image is summarized by comparing the pixels with their corresponding neighbors [7, 8]. Consider a pixel. It's surrounded by eight other pixels as neighbors. Now comparing the intensity of the center pixel with that of the neighbors if the intensity of the center pixel is higher than or equal to that of the neighbor under consideration then it's denoted by 1, else with 0. Thus, for each pixel a binary representation is formed, which leads to a total of 2^8 possible combinations, these combinations are termed as Local Binary Patterns. The OpenCV library of Python features a rich variety of face recognition algorithms through its FaceRecognizer class. The LBP algorithm is enabled by using the command createLBPHFaceRecognizer(). This system uses the LBP algorithm paired with the Yale Facedatabase or Yalefaces [6, 9] (Fig. 5).

3.5 Doorlock Module

The door lock module of this system is simulated using a DC motor to demonstrate the locking and unlocking function. This module is a combination of a relay driver circuit and a DC motor; this system uses an HFD27 Series 5V 1A 125 Ω (DPDT) Through Hole SubMiniature DIP Relay to control the DC motor. The driver circuit is also provided with leads for a 9 V battery to drive the motor when triggered. The driver is triggered by the core module through the GPIO pins.

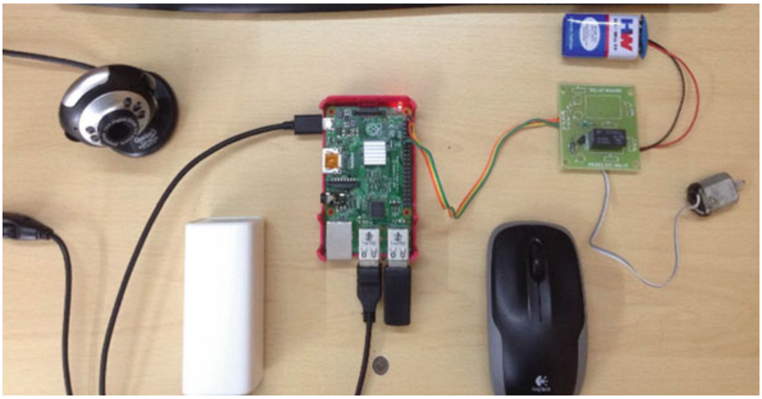


Fig. 6 Intial setup

4 Design and Implementation

This section emphasizes on the actual hardware implementation of the proposed system, the various modules, components, peripherals and the interconnections between them are discussed here. The first stage of the implementation is to prepare the Raspberry Pi 2 module for its first boot; this is done by downloading the latest version of the Raspbian operating system from the official Raspberry Pi website. A microSD card is the formatted using SD Formatter; it's then flashed with the Raspbian OS using Win32 Disk Imager [10]. The first boot is then completed on the Raspberry Pi connecting the required peripherals, such as power supply, keyboard, mouse, Ethernet cable, etc. (Fig. 6).

The Raspberry Pi for optimal operation requires a quality power supply; the Pi can be driven by using any Micro USB based mobile phone chargers with a good current rating, and this system is powered by a 5V 2A power bank for uninterrupted operation. Since the Raspberry Pi doesn't natively support wireless internet a USB WiFi dongle is used for connectivity; the Pi also has an Ethernet port which can be used to gain wired internet access [9]. Using Python programming language preinstalled on Raspbian the source code of the system is provided and tested appropriately. The USB Camera is interfaced, the GPIO pins are programmed using commands in Linux and Python in this stage. The camera is interfaced to the Pi via the USB port and the door lock module is interfaced via the GPIO pins on the Pi (Figs. 7 and 8).

5 Experimental Results

This section emphasizes on the final results of the proposed system, the system has been configured to recognize one of the author's face, and thus some face images

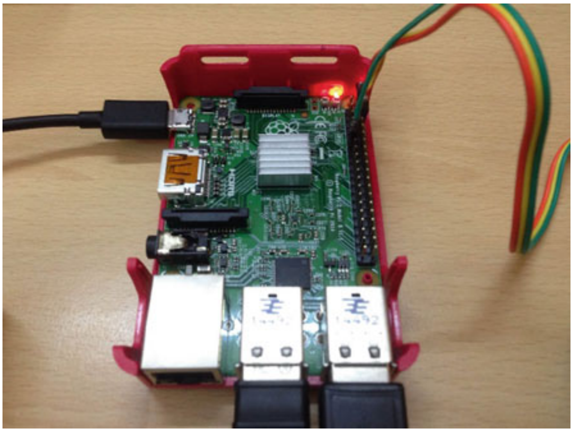


Fig. 7 GPIO and peripherals

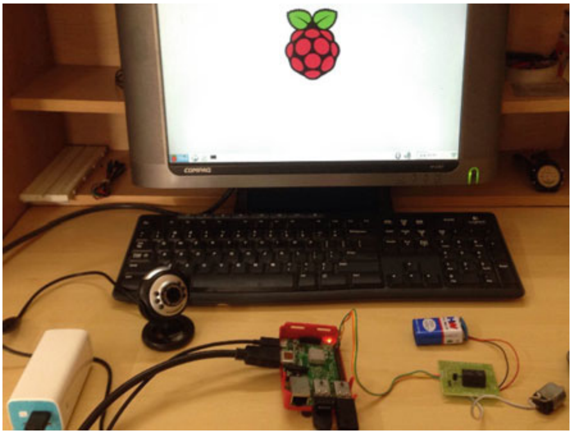


Fig. 8 Final setup

are taken in varying light conditions and are added to the database which is already populated with faces from Yale database. The system correctly recognizes the face and unlocks the door which is simulated by the DC Motor along with an SMS alert. In the case of an unauthorized person, the algorithm reports non-availability of the face in the database to the core module, which in turn forwards the live snapshot to the owner's email address for manual authentication along with an SMS alert. The owner now has to manually provide access to the person if he/she recognizes the person, this is done by replying to the Pi's email with a secure code as its subject, this code can be changed by the owner. Once the Pi receives this code, it validates it and unlocks the door.

6 Conclusion

This paper presents the design and implementation of an intelligent home security system using a robust, low-cost, low power single chip approach with the Internet as its backbone. This paper also explores the immense potential of computer vision in general and face recognition, in particular, the possibilities of IoT in home security and automation. The versatility and prowess of Linux operating system, the Python programming language, and the OpenCV library have also been explored, in depth.

References

1. Ahonen T, Pietikäinen M, Hadid M, Mäenpaä T (2004) Face recognition based on the appearance of local region. Machine Vision Group, InfoTech., University of Oulu, IEEE, Finland
2. Ahonen T, Hadid A, Pietikäinen M (2006) Face description with local binary patterns: application to face recognition. Draft
3. Faizi A (2008) Robust face detection using template matching algorithm. University of Toronto, Canada
4. Feng P (2004) Face recognition based on elastic template. Beijing University of Technology, China. Yang MH, Kriegman DJ, Ahuja N (2002) Detecting faces in images: a survey. IEEE Trans PAMI
5. Hadid A, Heikkilä M, Ahonen T, Pietikäinen M (2004) A novel approach to access control based on face recognition. Machine Vision Group, InfoTech Oulu and Department of Electrical and Information Engineering, University of Oulu, Finland
6. Rodriguez Y (2006) Face detection and verification using local binary patterns. Ph.D. thesis, Acole Polytechnique Federale de Lausanne
7. Nosaka R, Ohkawa Y, Fukui K (2012) Feature extraction based on co-occurrence of adjacent local binary patterns. In: Proceedings of the 5th Pacific Rim conference on advances in image and video technology, vol Part II, PSIVT 2011, pp 82–91
8. Zhang C, Zhang Z (2009) A survey of recent advances in face detection. In: Face recognition: face in video, age in variance, and facial marks. 2010 Unsang Park, Michigan State University
9. Zhang H, Zhao D (2004) Spatial histogram features for face detection in color images. In: IEEE 5th Pacific Rim conference on multimedia. Tokyo, Japan, pp 377–384
10. Brubaker S, Wu J, Sun J, Mullin M, Rehg J (2005) On the design of cascades of boosted ensembles for face detection. Technical report GIT-GVU-05-28, Georgia Institute of Technology

MediCloud: Cloud-Based Solution to Patient's Medical Records

G. B. Praveen, Anita Agrawal, Jainam Shah and Amalin Prince

Abstract Cloud computing has lately emerged as a new standard for hosting and delivering services over the internet. As per the Code of Federal Regulations, the hospitals are required to retain the radiological records for a duration of 5 years. Healthcare organizations are considering cloud computing as an attractive option for managing radiological imaging data. In this paper, we have proposed a public cloud-based "Infrastructure as a Service (IaaS)" model as a solution for maintaining the medical records of patients. New patient information is registered at the hospital with the help of a unique identification number. For each user, the bucket is created in the Amazon AWS cloud to store or retrieve the data. Data access is performed with the help of username and password provided by the web link which is embedded in the unique QR code. Two QR codes are used, the first code gives the access to the login page, whereas the latter one is used for accessing the corresponding user bucket.

1 Introduction

A prototype with a universal, on-demand access to a shared pool of resources which can be procured and supplied with nominal management effort or service provider interaction is defined as "Cloud Computing" [1, 2].

G. B. Praveen (✉) · A. Agrawal · J. Shah · A. Prince
Department of Electrical and Electronics Engineering, BITS Pilani – K.K Birla Goa Campus, Goa, India
e-mail: p20120404@goa.bits-pilani.ac.in

A. Agrawal
e-mail: aagrawal@goa.bits-pilani.ac.in

J. Shah
e-mail: f2014023@goa.bits-pilani.ac.in

A. Prince
e-mail: amalinprince@goa.bits-pilani.ac.in

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_105

Hospitals in the developing countries face funds shortage to invest in operation expansion. The healthcare industry is gradually adopting latest technology for storing digital images in the cloud database. Currently, the technology adoption is limited only to billing automation or back-office systems [3]. Digitization of medical records of patients is a vital and essential stage for quick and reliable storage and retrieval of patient's medical records [4]. However, it is not easy to maintain huge data centers, since heavy investment is required in human resources and technology for the same. The importance of cloud-based system is that data can be made accessed at any location and at any time. Cloud-based service provides pay as per use model so that massive investment such as buying and deploying expensive technologies can be avoided.

Rolim et al. [5] proposed an automated framework with the help of sensors to acquire vital data from the human body and to store the information in the cloud. A platform based on cloud computing framework is proposed for the management of mobile and wearable healthcare sensors [6]. Chang et al. [7] developed a cloud computing adoption framework (CCAF) and cloud storage design and deployment framework based on storage area network (SAN).

Client platform security is the major drawback of the current e-health solutions and standards. To overcome the security issue Lohr et al. [8] presented a secured framework for establishing privacy domains in e-health infrastructures. He et al. [9] proposed a private cloud platform architecture which consists of six layers and utilizes the message queue as a cloud engine. To overcome the drawbacks of the abovementioned frameworks, we propose a public cloud-based Infrastructure as a Service (IaaS) model as a solution for maintaining the medical records of patients. Characteristics and model deployment strategies of cloud computing are discussed in Sect. 2 and real-time challenges faced by the hospitals is described in Sect. 3. Section 4 describes the system model and proposed methodology is briefed in Sect. 5, concluding remarks are drawn in Sect. 6.

2 Characteristics and Deployment Models of Cloud Computing

The NIST has described the characteristics of the cloud as follows [1, 10, 11]:

- On-demand self-service: A user can demand computing capabilities, such as network storage, server time, etc., whenever required.
- Broad network access: Users can make use of the facilities over the network and access it through varied client platforms such as mobile phones, laptops, and tablets.
- Resource pooling: The different virtual and physical resources of the service provider are allocated and reallocated dynamically as per the customer demands. The computing resources are combined to serve numerous customers.

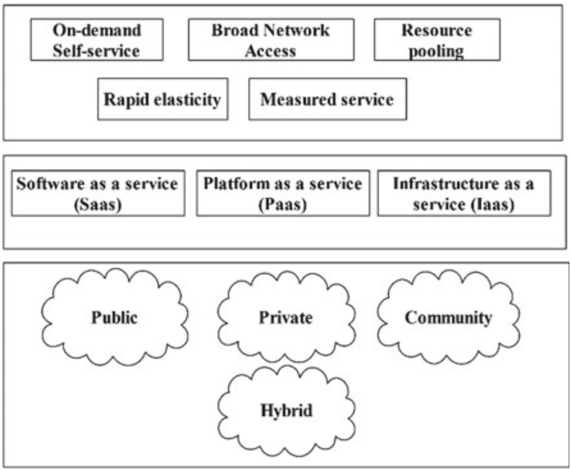


Fig. 1 Cloud deployment models

- Rapid elasticity: The computing capabilities often seem to be infinite and can be procured in required quantity as per the requirement. Capabilities can be demanded and released elastically. This process is generally automatic and it scales inward and outward proportionally with the demand.
- Measured service: Cloud systems provide a metering ability to control and optimize the usage of the resource.

The NIST has listed four categories of cloud deployment models as shown in Fig. 1 [1, 10–12]:

- Public cloud: The general public can use the services and resources via the internet. These services are typically low cost. Public cloud can be utilized by the general public, small and medium enterprise clients and even by the large organizations. Amazon AWS, Microsoft Azure, Google Cloud Platform are some of the providers of public cloud service.
- Private cloud: The cloud infrastructure is set up exclusively for an organization, although it may be managed by a third party or the organization itself. Management by third-party permits better controls over the infrastructure, security and management issues related to the cloud resources. The people outside the organization do not have access to it as it is restricted. It can be on the premises of the organization. This has less elasticity when compared with other deployment models.
- Community cloud: A group of organizations which share mutual interests and use the services. These services support a specific community which has similar concerns. Government institutions can share or members or partners of a common group can use community cloud.
- Hybrid cloud: It is a blend of two or more clouds (private, community, or public). This provides the most flexibility because it takes the benefits of all the other categories public, private, or community.

NIST definition of Cloud:

- Three service models of cloud computing was introduced by NIST [10, 13, 14]:
- Infrastructure as a Service (IaaS): The customer can demand processing capabilities, storage facilities, network abilities and also other fundamental computing resources which include operating systems, applications, and database systems. It is divided into compute clouds and resource clouds. Users can demand and access computational resources such as RAM, memory, CPU, processing power, virtual machines in compute cloud whereas, in resource cloud, users can access enhanced virtualization capabilities. They provide scalable resources as services to users.
- Platform as a Service (PaaS): The consumers can develop and host applications and software on a platform which provides computational resources. The behavior of the server can be controlled by making use of dedicated application programming interfaces (APIs) according to user requests.
- Software as a Service (SaaS): The provider's applications which run on the cloud are provided as a software service to the customers. The consumers can access the hosted applications or software over the internet.

3 Challenges

There are various challenges to the hospitals which are directly related to their Information Technology (IT) department.

3.1 *Data Storage*

The volume of digital data that is generated by healthcare systems has grown exponentially in the past decade. It is not an easy job to maintain this data. The medical regulations of country specify for how long the data must be stored and archived. For example, healthcare providers need to retain medical records archived for 6 years after discharge according to Health Insurance Portability and Accountability Act (HIPAA) [15]. Data backup and recovery has to be done as a measure to protect the data.

3.2 *Data Sharing*

Patients need to carry hard copies of scans and reports every time they visit a doctor. The data might be digitized but even though sharing and accessing it from any location is difficult.

3.3 Security

Patient's image scans and reports need to be maintained confidentially [16]. Secured data storage is an important task, hence these data should be encrypted and saved in the cloud. Encryption Algorithm implementation is a difficult task for the hospital's IT department.

3.4 Availability and Data Corruption

Availability of a system is a measure of how reliable the system is. The data must be available as and when required. The data stored may get corrupted over time due to bit-flip errors.

3.5 High Cost of Maintaining Infrastructure

It is not easy to maintain huge data centers. It requires a huge investment in human resources and technology [3]. Also, the current systems need upgradation which requires capital.

4 System Model

This section deals with the definition of our proposed model which consists of the below-mentioned entities.

- Object: An object consists of any files of types (e.g., PDF, JPEG, DOC, etc.) [17].
- Bucket: Buckets are the containers for objects [17].
- Identity and Access Management services (IAM): AWS identity and web management is a web service by which the AWS resource access can be controlled. The identity management portion of the IAM service is required for user authentication (sign in) and authorization. The access management portion of the IAM service defines the permission for a user or other entity in an account. Permissions are granted through policies that are created and then attached to users [18].
- Registered user: The patient who comes to a hospital and gets registered is called a registered user.
- Root user: The admin of the system is called as the root user. The root user has complete access to all the services and resources.

5 Methodology

The proposed framework uses public cloud such as Amazon AWS for maintaining the medical records of patients as depicted in Fig. 2.

When the patient is first registered at the hospital (registered user), the patient will be given a unique identification number which will be used to identify each patient uniquely. A bucket will be created for that registered user by the root user. Access Control Policy will be specified such that only the bucket corresponding to that registered user can be accessed by that registered user. No registered user can access any other registered user's bucket. For this purpose, identity access and management (IAM) services are used. Username and password are generated corresponding to each registered user. QR code which contains the web link for accessing the data will be given. QR code is generated by a Ruby Script which uses Barby gem. On scanning the QR code, the user will be redirected to a login page. Another QR code, which contains the username, password, and the steps to be followed will be given. On scanning this QR code, the text will appear. The user has to enter the username and password on the login page and then follow the steps that appeared in the QR code as per the requirements, i.e., to store new data or to retrieve previously stored data. QR code is printed and given so that the user does not have to remember the

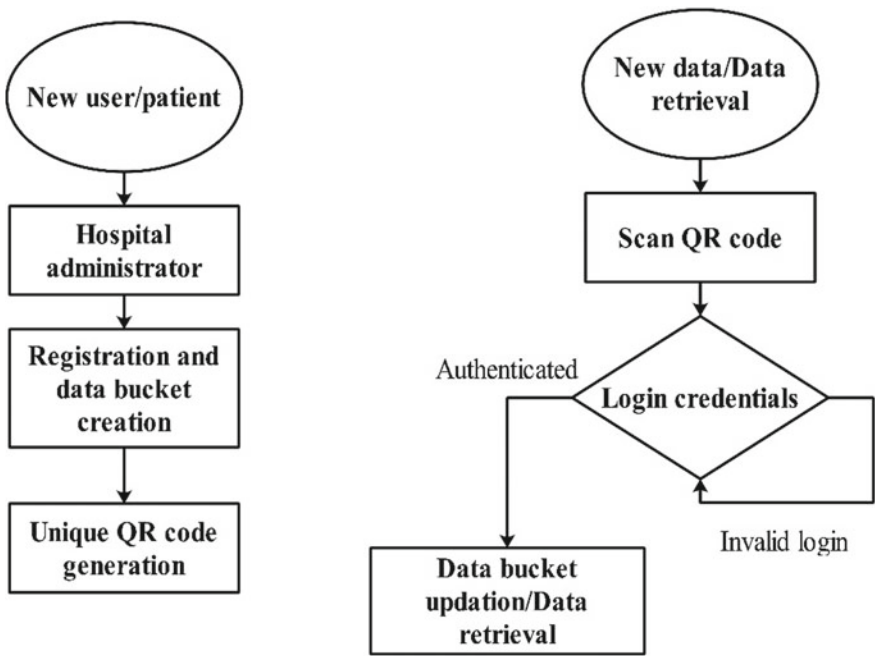


Fig. 2 Proposed methodology

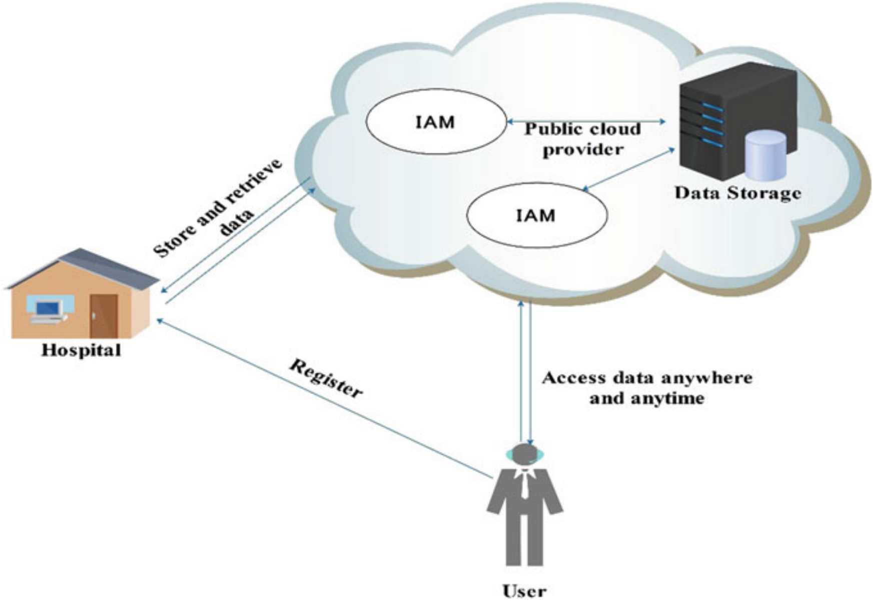


Fig. 3 Overview of medical cloud

links, username, password and the steps to be followed. A complete overview of the proposed system is shown in Fig. 3.

The abovementioned methodology solves the challenges listed in the previous section.

5.1 Durability

The objects are stored on multiple devices across multiple facilities by Amazon Simple Storage Service (S3) in Amazon S3 region redundantly. The service is designed in such a way that concurrent device failures are quickly detected and any lost redundancy is repaired. When a data storage request is received, the data gets stored in multiple data centers [19].

5.2 Accessibility

The data to be retrieved is accessible from any location anytime. It can be accessed on a computer, laptops, and even on smart mobile phones.

5.3 Data Encryption

Users are authenticated before access to the data. The data uploaded and downloaded to the cloud service provider is secured via SSL endpoints using the HTTPS protocol. The security of data stored in the cloud when at rest is achieved by using the security features provided by the cloud provider. Amazon S3 provides server-side encryption for the data at rest. The data is encrypted by one of the strongest block ciphers available, i.e., the AES-256 (Advanced Encryption Standard) [19].

5.4 Availability and Detection of Data Corruption

Amazon S3 checks for data corruption by applying algorithms such as a combination of MD5 checksums and cyclic redundancy checks (CRCs). These checksums are performed on data at rest and any corruption found is repaired using redundant data [19].

5.5 Scalability

Cloud provides the ability of scalability [20]. There is no need to guess the amount of data that will be generated as is the case with data centers. The storage size can be increased and decreased as and when required. The pay as per use model of cloud is very beneficial as you need to pay only for the services used.

6 Conclusion

Robust data management of patient information is facilitated by the cloud storage system. The image scans and reports of the patients get stored in the cloud database, which facilitates data sharing across any location and at any time. In this paper, we have proposed a public cloud-based infrastructure as a service (IaaS) model as a solution for maintaining the medical records of patients. Two QR codes are generated for the accessibility of the login page and bucket access respectively. Cloud storage provides a scalable environment for businesses to increase or decrease storage on-demand. Cloud storage provides durability since data is stored in multiple data centers, and failure of one data center will not affect the stored data in the other center. Cloud storage has transformed the capital-intensive system and saves costs due to its pay as per usage model. Secured data storage and reduction of operating risks are the main advantages of using cloud storage system. The healthcare organizations such as hospitals need cost-effective and innovative methods to deal with the challenges mentioned in the previous sections. We intend to apply the solution proposed above to real-world application in future works in order to assess the advantages of the proposed solution.

Appendix

A. QR Code

QR code is the acronym for the Quick Response code. QR Code is a kind of 2D barcode. Large capacity, Small printout size, high-speed scanning, etc., are some of the features of QR code [21]. It can encode up to 4296 alphanumeric characters in a single pattern [22]. In a QR code, there are three large square patterns (there is a small black square surrounded by white bars inside it). There is an additional square in newer versions. The other area of QR code is used to encode the embedded information and it consists of a number of small blocks. URLs and alphanumeric characters can be easily encoded in a QR code. Easy access to the website is possible by scanning the QR code. QR code scanning just requires a mobile phone with a camera.

B. AES-256 (Advanced Encryption Standard)

AES is a modern block cipher which supports key lengths of 128, 192, and 256 bits. As the name suggests, AES-256 uses a key length of 256 bits. AES is a symmetric key algorithm which uses the same key for encryption and decryption. The best-known attack till date against the AES is the brute force attack [23]. Practically, no one without the key can read the data encrypted by AES.

C. Cyclic Redundancy Checks (CRCs)

Cyclic Redundancy check is an error detecting code. Accidental changes to data caused by bit-flip errors can be detected by this algorithm. When the blocks of data enter the system, a short check value is attached which is calculated on the basis of the remainder of a polynomial division of the contents of data. The calculation is again done when retrieving data. If both the values do not match, then data is found to be corrupted. Thus, corrective measures can be taken. CRCs are so called because the algorithm is based on a cyclic code and the data verification value (check value) is a redundancy (it expands the message without adding information) [24].

References

1. Mell P, Grance T (2011) The NIST definition of cloud computing
2. Mell P, Grance T (2010) The NIST definition of cloud computing. Commun ACM 53(6):50
3. Chand R, Tripathi M, Mishra SK (2016) Cloud computing for medical applications and health-care delivery: technology, application, security and swot analysis

4. Aziz HA, Guled A (2016) Cloud computing and healthcare services
5. Rolim, CO et al (2010) A cloud computing solution for patient's data collection in healthcare institutions. In: Second IEEE international conference on eHealth, telemedicine, and social medicine, pp 95–99
6. Doukas C, Maglogiannis I (2012) Bringing IoT and cloud computing towards pervasive healthcare. In: Sixth IEEE international conference on innovative mobile and internet services in ubiquitous computing (IMIS), pp 922–926
7. Chang V, Walters RJ, Wills G (2012) Cloud storage and bioinformatics in a private cloud deployment: lessons for data-intensive research. In: International conference on cloud computing and services science. Springer, pp 245–264
8. Lohr H, Sadeghi A-R, Winandy M (2010) Securing the e-health cloud. In: Proceedings of the 1st ACM international health informatics symposium, pp 220–229
9. He C, Fan X, Li Y (2013) Toward ubiquitous healthcare services with a novel efficient cloud platform. IEEE Trans Biomed Eng 60(1):230–234
10. National Institute of Standards and Technology (2009). Retrieved from www.nist.gov
11. Wang L, Alexander CA (2013) Medical applications, and healthcare based on cloud computing. Int J Cloud Comput Serv Sci 2(4):217
12. Shadi JH, Tim B, Wei G, Dachuan H, Song W (2010) Cloud types and services. In: Handbook of cloud computing. Springer, pp 335–355
13. Vilaplana J, Solsona F, Abella F, Filgueira R, Rius J (2013) The cloud paradigm applied to e-Health. BMC Med Inform Decis Mak 35(13):1–10
14. Chang V, Roure DD, Wills G, Walters RJ, Barry T (2011) Organizational sustainability modeling for return on investment (ROI): case studies presented by a National Health Service (NHS) Trust UK. CIT J Comput Inf Technol 19(3):177–192
15. John N, Shenoy S (2014) Health cloud-healthcare as a service (HaaS). In: IEEE International conference on advances in computing, communications, and informatics (ICACCI), pp 1963–1966
16. Rolim CO, Koch FL, Westphal CB, Werner J, Fracalossi A, Salvador GS (2010) A cloud computing solution for patient's data collection in healthcare institutions. In: Second IEEE international conference on eHealth, telemedicine, and social medicine, pp 95–99
17. Amazon Simple Storage Service (S3) Introduction. Retrieved from http://docs.aws.amazon.com/AmazonS3/latest/dev/Introduction.html
18. Amazon Identity and Access Management (IAM) Services Introduction. Retrieved from http://docs.aws.amazon.com/IAM/latest/UserGuide/introduction.html
19. Amazon Simple Storage Service. Retrieved from https://aws.amazon.com/s3/faqs/
20. Marwan M, Kartit A, Ouahmane H (2016) Cloud-based medical image issues. Int J Appl Eng Res 11(5):3713–3719
21. Sun A, Sun Y, Liu C (2007) The QR-code reorganization in illegible snapshots taken by mobile phones. In: IEEE international conference on computational science and its applications (ICCSA), pp 532–538
22. Kan TW, Teng CH, Chou WS (2009) Applying QR code in augmented reality applications. In: ACM proceedings of the 8th international conference on virtual reality continuum and its applications in industry, pp 253–257
23. Paar C, Pelzl J (2009) Understanding cryptography: a textbook for students and practitioners. Springer, Berlin
24. Sarwate DV (1988) Computation of cyclic redundancy checks via table look-up. Commun ACM 31(8):1008–1013

A Trio Approach Satisfying CIA Triad for Medical Image Security

Sivasaranyan Guhan, Sridevi Arumugham, Siva Janakiraman, Amirtharajan Rengarajan and Sundararaman Rajagopalan

Abstract Medical image security attains a great appeal due to the challenges in transmitting them through an open network channel. Even a small change in medical information leads to a wrong diagnosis. Hence, methods to prevent the attacks and tampering on the medical images in the open channel have a great demand. This work proposes such an attractor-assisted medical image watermarking and double chaotic encryption which preserves the confidentiality, integrity and authenticity of the medical image. Patient diagnosis details are compressed using lossless compression approach and embedded in the pixels of patient's DICOM image. Attractor-based selective watermarking is implemented using integer wavelet transform. Encryption will be carried out by employing chaotic maps using confusion and diffusion operations. The effectiveness of the work will be evaluated using standard analyses namely MSE, PSNR, SSIM, entropy, correlation, histogram and key space.

1 Introduction

Recent trends in technological growth are exponential. This exponential growth led to multimedia communication in medicine industry for telemedicine. Multimedia communication in telemedicine is established by the use of electronic health record technology. Electronic health records (EHRs) like health history data of a patient, demographic data, physical examination data, laboratory test solutions, treatment measures and prescriptions are extremely confidential in nature. Information security is achieved by ensuring that the (i) transmitted medical images cannot be retrieved by unauthorized parties (Confidentiality), (ii) received images are not tampered during transmission (Integrity), and (iii) images are from original sources and reach the original receivers (Authentication).

S. Guhan · S. Arumugham · S. Janakiraman · A. Rengarajan · S. Rajagopalan (✉)
School of Electrical and Electronics Engineering, SASTRA Deemed to be University, Thanjavur 613401, India
e-mail: raman@ece.sastra.edu

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_106

In this paper, we have proposed a hybrid model which includes three major security techniques like encoding, watermarking and encryption that ensures the aforementioned prerequisites. Many compression techniques have been used worldwide to reduce size and traffic. Using lossless compression, the received image is as same as the transmitted image that is, no information is lost [1]. Many previous researches have shown that arithmetic encoding is the most optimum technique and achieves higher compression techniques than other lossless techniques. In this paper, arithmetic encoding is used to encode the image. Arithmetic encoding is variable length encoding technique that achieves lossless compression [2]. Arithmetic encoding encodes the target image into binary bit stream.

Many researchers have proposed many digital watermarking techniques. Kavitha and Shan [3] proposed a watermarking scheme for DICOM image using IWT. Alpha blending watermarking requires the receiver to have the original host image. This can be disadvantageous in most cases. Hence, a lot of research is happening in the bit plane watermarking [4]. Watermarking can be classified into two major areas. Many researches on watermarking suggest that frequency domain watermarking provides higher robustness and imperceptibility [5]. Due to the rapid development of the Internet, the need for digital product protection is on the rise. This concern led to the need for more secure digital watermarking technologies [6]. Recent researches in frequency domain watermarking show that IWT is computationally faster and provides better results than DWT. Despite many researches in this technology, very few researchers have proposed selective watermarking. Singh et al. [7] proposed multiple watermarking algorithms by using selective DWT coefficients for embedding. Malonia and Agarwal [8] elucidated an algorithm where watermarking is done on the DWT coefficients selected by arithmetic progression. The results from selective embedding show that the watermark has higher robustness and imperceptibility.

Hybridization of watermarking with encryption has proved that it can significantly increase in the medical image security [9]. Though there are many algorithms available for encryption, chaotic maps have drawn a lot of attention. Because of the nonlinearity and unpredictability, the chaotic maps provide researchers have focussed on using it for the encryption purpose. Safi and Maghari [10] examined the competence of a combination of two chaotic maps. The results are better and less complicated than many other proposals with similar results. Shyamala [11] proposed chaotic confusion—diffusion algorithm. It makes use of the fact that mathematical equations generated by chaotic maps are reversible and efficient enough to cause diffusion on the image pixels. Huang and Yang [12] introduced two encryption algorithm to randomize transforms, chaotic scrambling, chaotic permutation and chaotic diffusion. Wang et al. [13] proposed an image encryption using a combination of 1D chaotic maps to confuse the image. The experimental results showed that the scheme is resistive to different attacks.

2 Preliminaries

2.1 Lorenz Attractor (LA)

A system may tend to develop towards a set of numerical values for a wide variety of initial conditions. These set of numerical values are called attractors. Lorenz attractor is one of the attractors that provides a chaotic solution. Lorenz attractor is a set of chaotic solutions of the Lorenz system. Lorenz system is modelled by three ordinary differential equations. For system that behaves chaotic when the constants $\sigma = 10$; $\rho = 28$; $\beta = 8/3$ [14].

$$\frac{\mathrm{d}x}{\mathrm{d}t} = \sigma(y - x) \tag{1}$$

$$\frac{\mathrm{d}y}{\mathrm{d}t} = x(\rho - z) - y \tag{2}$$

$$\frac{\mathrm{d}z}{\mathrm{d}t} = (xy) - (\beta z) \tag{3}$$

2.2 Logistic Map

Chaotic maps are functions that are highly sensitive to initial conditions and exhibit chaotic behaviour. Logistic map is one such 1D map which exhibits chaotic behaviour from a simple nonlinear dynamic equation. Logistic map is given by

$$x_{n+1} = r x_n (1 - x) \tag{4}$$

x_n should be a number between 0 and 1. Logistic map becomes chaotic when the constant r ~ 3.5699. For different constant values and seed value x_0, the map shows different characteristics.

2.3 Tent Map

Tent map is another such 1D chaotic map that exhibits chaotic behaviour from a simple nonlinear dynamic equation. Tent map is given by

$$x_{n+1} = \mu x_n \quad \text{for } x_n < 1/2 \tag{5}$$

$$x_{n+1} = \mu x_n (1 - x_n) \quad \text{for } x_n \geq 1/2 \tag{6}$$

x_n should be a number between 0 and 1. When constant $\mu = 2$, the system maps the interval [0, 1] onto itself. The periodic points are dense in [0, 1], so the map becomes chaotic. For different constant values and seed value x_0, the map shows different characteristics.

2.4 Integer Wavelet Transform

Integer wavelet transform is used for its finite integer coefficients which makes the transform lossless. Due to the fact that all the coefficients of IWT are integers, there is no rounding off errors so the watermark can be extracted without any loss. Lifting scheme is implemented to perform IWT. Lifting scheme consists of three stages. They are split, predict and update.

Split: The main signal is decomposed into even sequence and odd sequence.
Predict: The numbers from one sequence are predicted on the basis of other sequence.
Update: Even samples are updated from the input even samples and the updated odd samples.

3 Proposed Work

The proposed method introduces randomness while embedding and encrypts the watermarked image taking advantage of various chaotic maps. In this work, a medical EHR 1 of size 128 × 128 is taken as secret image and a host 8 bit DICOM of size 512 × 512 is taken for validations. The use of Lorenz attractors, logistic maps and tent map for watermarking and encryption is explained in the detailed block diagram Fig. 1.

3.1 Watermarking

i. Secret image is encoded using arithmetic encoding technique which gives binary bit stream of length N.
ii. Permutation operation is performed when binary sequence is left cyclic shifted by a seed value.
iii. The bit stream is divided into three arrays such that $(1 - i)$, $((i + 1) - j)$ and $((j + 1) - N)$ where $1 \leq i, j \leq N$ and $i \neq j$.
iv. Host image of size $M \times N$ is decomposed using integer wavelet transform into its sub-bands LL, LH, HL and HH each of size $M/2 \times N/2$.
v. The HH sub-band is subdivided into three segments say HH_1, HH_2 and HH_3.

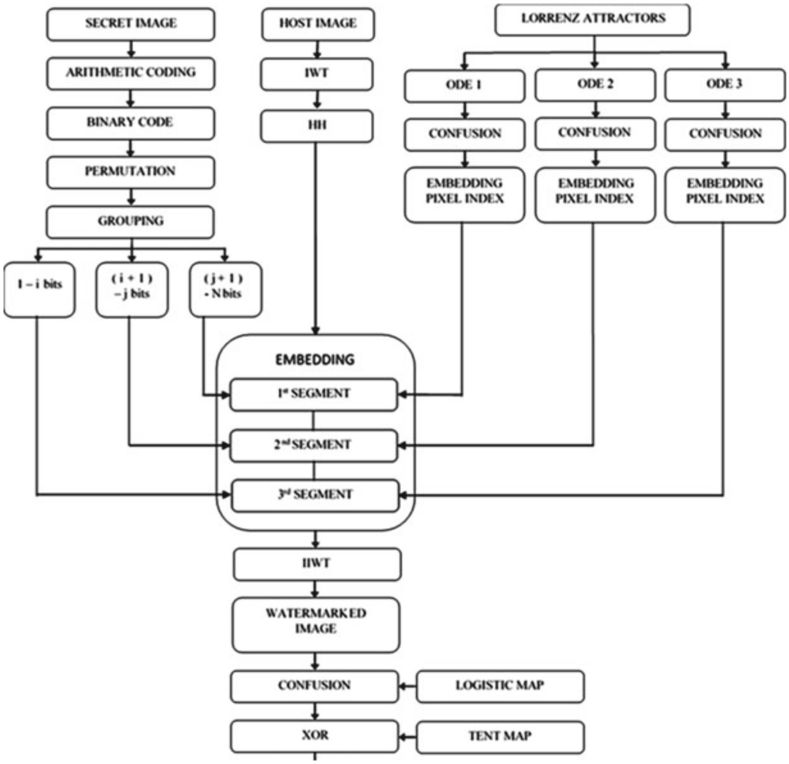


Fig. 1 Detailed block diagram of the proposed method

vi. Lorenz attractor (LA) ordinary differential Eqs. (1)–(3) are used to generate pseudo random number sequence

$$S_{\mathrm{LA1}} = \left(S_1, S_2, \ldots, S_{M/6\times N/6}\right)$$
$$S_{\mathrm{LA2}} = \left(S_{M/6+1}, S_{M/6+2}, \ldots, S_{M/3\times N/3}\right)$$
$$S_{\mathrm{LA3}} = \left(S_{M/3+1}, S_{M/3+2}, \ldots, S_{M/2\times N/2}\right)$$

vii. Perform sorting in ascending order on the pseudo random sequence $S_{\mathrm{LA1}}, S_{\mathrm{LA2}}$ and S_{LA3} as follows:

$$[I_{s1}, Y_{s1}] = \mathrm{sort}(S_{\mathrm{LA1}})$$
$$[I_{s2}, Y_{s2}] = \mathrm{sort}(S_{\mathrm{LA2}})$$
$$[I_{s3}, Y_{s3}] = \mathrm{sort}(S_{\mathrm{LA3}})$$

Where [•,•] = sort (•) is the sequence indexing function; I_s is the new index after sorting the data S and Y_s is the sorted sequence after sorting S.

viii. Each secret array is embedded into the first bit plane of the coefficients in the I_s position of each segments.

ix. Perform inverse integer wavelet transform on the LL, LH, HL and modified HH sub-bands to get the final watermarked image of size $M \times N$.

3.2 Encryption

i. Logistic chaotic map Eq. (4) is used to generate pseudo random number sequence

$$S_{\text{logistic}} = (S_1, S_2, \ldots, S_{M\times N})$$

which forms the key 1.

ii. Perform sorting in ascending order on the pseudo random sequence S_{logistic} such that

$$\left[I_{\text{logistic}}, Y_{\text{logistic}}\right] = \text{sort}\left(S_{\text{logistic}}\right)$$

where I_{logistic} is the new index after sorting the data S_{logistic} and Y_{logistic} is the sorted sequence after sorting S_{logistic}.

iii. Perform confusion on the pixels of watermarked image of size $(M \times N)$ such that

$$\text{Logistic}(i) \Rightarrow \text{Watermarked}\left(I_{\text{logistic}}, (i)\right),$$

where $1 \leq i \leq (M \times N)$

This results in first level chaotic encryption.

iv. Scale up $\text{Watermarked}_{\text{logistic}}$ by using, $X_{\text{scaled}} = \text{mod}(\text{Watermarked}_{\text{logistic}} \times 10^{14}, 256)$.

v. Tent chaotic map Eq. (5) is used to generate pseudo random number sequence

$$S_{\text{tent}} = (S_1, S_2, \ldots S_{M\times N})$$

which forms the key 2.

vi. Perform XOR to diffuse the confused image with key 2. This results in second level chaotic encryption and final encrypted image

$$\text{Image}_{\text{enc}} = X_{\text{scaled}} \oplus S_{\text{tent}}.$$

vii. Decryption and extraction are the reverse processes of encryption and watermarking, respectively.

4 Results and Discussion

The proposed work includes three modules namely compression, watermarking and encryption processes. The compression modules are validated by calculating compression ratio and average length. The watermarking module is validated by calculating MSE, PSNR and SSIM. The encryption module is validated by calculating correlation coefficients, entropy, histogram analysis, NPCR analysis and key space analysis. The simulation work of the proposed work is done using MATLAB R2016b in a system with 4 GB RAM, 500 GB Hard drive, Intel Core i3 processor and Windows 8.1 operating system. Figure 2 shows the final output image of each module.

The same secret image as shown in Fig. 2a is watermarked within different DICOM images and encrypted. The DICOM images and its corresponding encrypted outputs are shown in Fig. 3.

4.1 Compression Ratio

Compression Ratio is defined as ratio of number of bits in the original image and number of bits in the compressed image. Compression ratio is used to validate the efficiency of the used compression technique. Compression ratio is given by

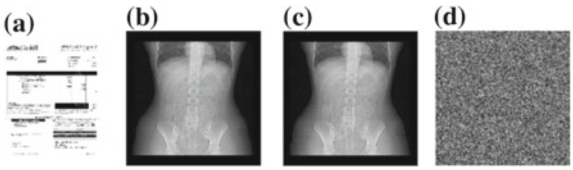


Fig. 2 **a** EHR 1, **b** original DICOM 1, **c** watermarked DICOM image, **d** encrypted image ENC 1

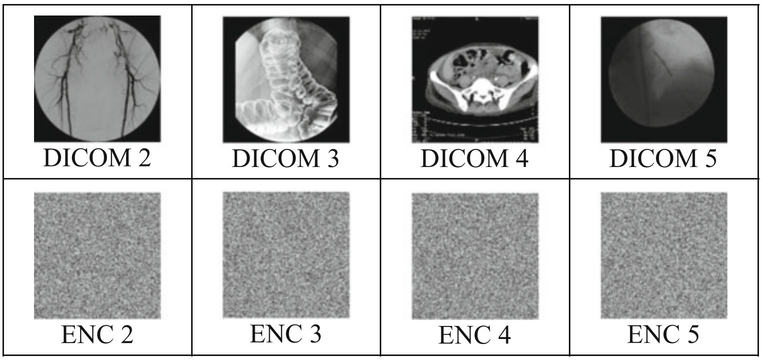


Fig. 3 Host DICOM images and its corresponding cipher images

Table 1 Compression ratio and average length

Image	Compression ratio	Average length
EHR 1	1.6753	4.7754
EHR 2	1.5039	5.3196
EHR 3	1.3930	5.7429
EHR 4	1.3905	5.7532
EHR 5	1.8143	4.4094

$$\text{Compression ratio} = \frac{\text{Number of bits in the original image}}{\text{Number of bits in the compressed image}}$$

4.2 Average Length

Average length is defined as the average of number of bits required to encode a symbol in a sequence. Arithmetic coding uses less number of bits for symbols with high probability. The compression ratio and average length of various secret images are validated and shown in Table 1.

4.3 Mean Square Error

Error metrics such are used to compare two image processing techniques. Mean square error is one such metric which is used to validate the proposed technique. Mean square error is the cumulative squared error between the original image and watermarked image. The ideal value of MSE between the two images is 0. Mean squared error is given by

$$\text{MSE} = \frac{1}{MN}\sum_{y=1}^{M}\sum_{x=1}^{N}\left[I(x,y) - I'(x,y)\right]^2$$

x Row
a Total number of rows
y Column
b Total number of columns.

Table 2 MSE, PSNR and SSIM values

Image	MSE	PSNR	SSIM
DICOM 1	0.0284	63.5999	0.8703
DICOM 2	0.0281	63.6386	0.9420
DICOM 3	0.0281	63.6504	0.9079
DICOM 4	0.0283	63.6122	0.9386
DICOM 5	0.0282	63.6251	0.8931

4.4 Peak Signal-to-Noise Ratio (PSNR)

Peak signal-to-noise ratio is the ratio of peak power of signal to power of noise in an image. PSNR is calculated between the watermarked image and the original image. The ideal value of PSNR is infinity. The mathematical formula for PSNR is

$$\mathrm{PSNR} = 20 * \log_{10}\left(\frac{\mathrm{MAX}}{\sqrt{(\mathrm{MSE})}}\right)$$

MAX Maximum possible pixel value of the image.

4.5 Structural Similarity Index

Structural similarity index is a full reference index used to measure the similarity between two images. The ideal value of SSIM is 1. SSIM is measured between two images of same size using

$$\mathrm{SSIM} = \frac{\left(2\mu_x\mu_y + c_1\right)\left(2\sigma_{xy} + c_2\right)}{\left(\mu_x^2 + \mu_y^2 + c_1\right)\left(\sigma_x^2 + \sigma_y^2 + c_2\right)}$$

μ_x average of x
σ_x variance of x
σ_{xy} covariance of x and y
μ_y average of y
σ_y variance of y
c_1, c_2 constants to stabilize the division.

From Table 2, the proposed method gives MSE nearly equal to 0, PSNR value is above the acceptable range of 40 dB and the SSIM value is nearly 1 showing that the proposed method is much better than the previous works.

Table 3 Correlation coefficients and entropy of DICOM images

Images	Entropy		Correlation coefficients			
	Original	Cipher	Direction	Original	Cipher	Ref. [15]
DICOM 1	6.4444	7.9907	H V D	0.9868 0.9892 0.9800	−0.0019 0.0040 0.0004	0.0014 −0.0009 0.0065
DICOM 2	4.9982	7.9778	H V D	0.9777 0.9835 0.9707	−0.0006 0.0024 0.0006	0.0081 −0.0039 0.0030
DICOM 3	5.7158	7.9536	H V D	0.9872 0.9859 0.9768	0.0032 −0.0017 0.0020	−0.0032 0.0016 −0.0029
DICOM 4	3.8657	7.9714	H V D	0.9131 0.9172 0.8713	−0.0007 −0.0005 0.0001	0.0075 0.0142 0.0061
DICOM 5	4.7728	7.9650	H V D	0.9972 0.9958 0.9962	0.0017 0.0024 −0.0027	NA

4.6 Entropy

Entropy measures the randomness among the pixels of an image. The ideal value of an 8-bit image is 8. The proposed method gives entropy value close to 8 as shown in Table 3.

$$H = -\sum_{i=1}^{n} P_i \times \log_2 P_i$$

H Entropy of the Image
N Grey level of an input image (0–255)
P_i Probability of the occurrence of symbol i.

4.7 Correlation Coefficient

Correlation coefficient is statistical metric used to validate the encrypted image. It is a numerical measure of relationship between two adjacent pixels of an image. For an original image, the coefficient is 1 which indicates that there is strong agreement between the pixels. And for an encrypted image, the value should be very less. Figure 4 shows the correlation graph for a test image. The statistical validation of the proposed method is proved from Table 3. The metrics are compared with an existing algorithm to prove its efficiency.

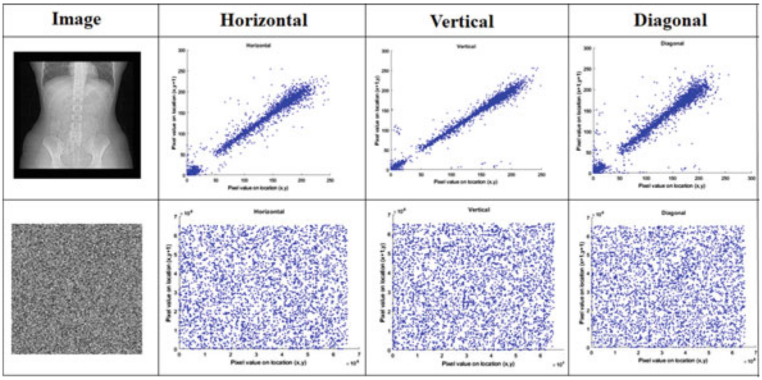
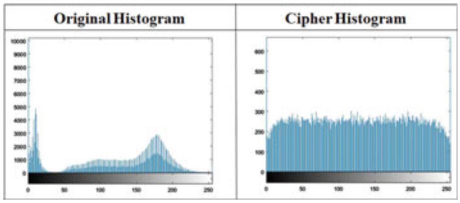


Fig. 4 Correlation graph of plain image and encrypted image

Fig. 5 Histogram analysis of plain and cipher image of DICOM 1



$$x = \frac{n\left(\sum ab\right) - \left(\sum a\right)\left(\sum b\right)}{\sqrt{\left[n\sum a^2 - \left(\sum a\right)^2\right]\left[n\sum b^2 - \left(\sum b\right)^2\right]}}$$

x	Correlation Coefficient
n	Total number of Pixels
a and b	Adjacent pixels
H	Horizontal
V	Vertical
D	Diagonal.

4.8 Histogram Analysis

Histogram is the accurate representation of distribution of pixels of an image. Histogram is used to measure the intensity levels and its frequency of an image. From Fig. 5, it is clear that the original image has different levels of intensity and the encrypted image obtained from the proposed method has a flat response.

Table 4 NPCR analysis

Image	NPCR	Ref. [15]
DICOM 1	99.9866	99.6240
DICOM 2	99.8497	99.2360
DICOM 3	99.7509	99.1382
DICOM 4	99.9943	99.2360
DICOM 5	99.7757	99.6076

4.9 NPCR Analysis

Number of pixel change rate (NPCR) is a measure used to test the influence of plain image over the encrypted image. A perfect encryption algorithm is highly resistive to differential attacks and possesses better diffusion property. These values are calculated for different test images and shown in Table 4.

4.10 Key Space Analysis

Key space analysis is to check the feasibility of brute force attack. When the cipher image is decrypted using all possible keys, it is called brute force attack. The key space should be large enough to withstand the brute force attack. For an encryption algorithm with 128-bit key, it requires 1.3736×10^{31} years to decrypt the algorithm. The secret keys used in this proposed method includes Lorenz attractors, logistic map, tent map and a seed value for cyclic shift. The total number of initial conditions required for Lorenz attractors (6), logistic map (2) and tent map (2) is 10. The seed value N used for cyclic shift operation also makes up for key space. Hence, the proposed method achieves a key space of $N \times 10^{170} > 2^{128}$ which is large enough to resist the brute force attack. Comparing to previous works [10] which has a key space of 2^{68} and [13] which has a key space of 2^{128} the proposed work is highly secure.

5 Conclusion

This paper is aimed to produce a comprehensive image hiding and image encryption taking advantage of all available algorithms. Based on the results obtained, it is clear that combining major image processing techniques like compression, watermarking and encryption to form a hybrid model gives more security than many previous models. This method provides a more secure and reliable means of communication in the telemedicine industry by ensuring the CIA triad.

References

1. Shahbahrami A, Bahrampour R, Rostami MS, Ayoubi M (2011) Evaluation of Huffman and arithmetic algorithms for multimedia compression standards. Int J Comput Sci Eng Appl 1(4):34–47
2. Mukesh PS, Pandya MS, Pathak S (2013) Enhancing AES algorithm with arithmetic coding. In: Proceedings of the 2013 international conference green computing, communication and conservation of energy, ICGCE 2013, pp 83–86
3. Kavitha KJ, Shan BP (2017) Implementation of DWM for medical images using IWT and QR code as a watermark. In: 2017 Conference on emerging devices and smart systems, ICEDSS 2017, Mar 2017, pp 252–255
4. Tyagi S, Singh HV, Agarwal R, Gangwar SK (2016) Digital watermarking techniques for security applications. In: 2016 International conference on emerging trends in electrical electronics and sustainable energy systems, pp 379–382
5. Zhang Y (2009) Digital Watermarking technology: a review. In: 2009 ETP international conference on future computer and communication, pp 250–252
6. Gupta P, Parmar G (2017) Image watermarking using IWT-SVD and its comparative analysis with DWT-SVD. In: 2017 International conference on computer, communication and electronics COMPTELIX 2017, pp 527–531
7. Singh AK, Kumar B, Dave M, Mohan A (2015) Multiple watermarking on medical images using selective discrete wavelet transform coefficients. J Med Imaging Health Inform 5(3):607–614
8. Malonia M, Agarwal SK (2016) Digital image watermarking using discrete wavelet transform and arithmetic progression technique. In: 2016 IEEE students' conference on electrical, electronics and computer science, pp 1–6
9. Aparna P, Kishore PVV (2018) An efficient medical image watermarking technique in e-healthcare application using hybridization of compression and cryptography algorithm. J Intell Syst 27(1):115–133
10. Safi HW, Maghari AY (2017) Image encryption using double chaotic logistic map. In: International conference on promising electronic technologies, pp 66–70
11. Shyamala N (2017) Reversible chaotic encryption techniques for images. pp 2–6
12. Huang H, Yang S (2016) Colour image encryption based on logistic mapping and double random-phase encoding. IET Image Process 11(4):211–216
13. Wang H, Xiao D, Chen X, Huang H (2018) Cryptanalysis and enhancements of image encryption using combination of the 1D chaotic map. Sig Process 144:444–452
14. https://en.wikipedia.org/w/index.php?title=Lorenz_system&oldid=833739483
15. Chandrasekaran J, Thiruvengadam SJ (2017) A hybrid chaotic and number theoretic approach for securing DICOM images. Secur Commun Netw 2017

A Novel Hybrid Method for Time Series Forecasting Using Soft Computing Approach

Arpita Sanghani, Nirav Bhatt and N. C. Chauhan

Abstract Improving the forecasting accuracy of time series is important and has always been a challenging research domain. From many decades, Auto-Regressive Integrated Moving Average (ARIMA) has been popularly used for statistic forecasting however it will solely forecast linear half accurately because it cannot capture the nonlinear patterns. Therefore here, we have projected a hybrid model of ARIMA and SVM. As Support Vector Machine (SVM) has demonstrated great outcomes in solving nonlinear regression estimation problems and to utilize the linear strength of ARIMA. Comparison with other models using different datasets has been done and the results are very promising.

1 Introduction

Time series prediction is an endlessly growing analysis domain. The accuracy of prediction is the main goal to realize. Prediction future is the best exploitation by time series prediction. With the statistic prediction, past data assortment of the constant variable is used to build a model to forecast long-term accessibility of those data. Then established model is utilized thus on extrapolate the statistic into the longer term. This modeling approach is very useful once little information is obtainable regarding past variables and no different things are thought [1–5].

A. Sanghani (✉)
Department of Computer Engineering, BVM Engineering College, Vallabh Vidyanagar 388120, Gujarat, India
e-mail: arpita.sanghani@gmail.com

N. Bhatt
Department of Information Technology, CSPIT, CHARUSAT, Anand 388421, Gujarat, India
e-mail: niravbhatt.it@charusat.ac.in

N. C. Chauhan
Department of Information Technology, A.D. Patel Institute of Technology, New V.V. Nagar, Anand 388121, Gujarat, India
e-mail: narendracchauhan@gmail.com

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_107

First, only the ARIMA model was used and was popular for time series forecasting but it was observed that individual models were not capable to handle nonlinearity. So, combine modeling came into the picture. Many combine models were introduced. And the results of some of them were also better than the individual. ARIMA works well for linear forecasting but it has shown a drawback of not giving good results when applied to nonlinear data. And after that ARIMA + ANN hybrid model came into the picture with various versions and it started giving goods results for the different type of time series data. But neural network also has some drawback and as we know that SVM is a good alternative to ANN so there is still scope for forecasting results improvement [6–8].

Our purpose in this paper is to do build a novel hybrid model for time series forecasting using ARIMA and a unique soft computing and Support Vector Machines (SVM) based on machine learning. The idea for exploiting SVM came from literature survey [9]. This technique is used to accurately forecast statistic knowledge and the processes of the system are usually nonlinear, nonstationary and not highlighted earlier. SVMs additionally verified to beat alternative nonlinear ways along with neural network primarily focused on nonlinear prediction ways like multilayer perceptrons. The detailed paper is organized as follows. In Sect. 1, introduction to recent time series trends, Sect. 2 describes the traditional model ARIMA of time series. Section 3 describes regarding soft computing approach model for time series Support Vector Machine (SVM). Section 4 gives detail about the proposed model. Section 5 shows the dataset and results then concludes this work by the conclusion section of the proposed work and future goals.

2 Auto-Regressive Integrated Moving Average (ARIMA) Model

For statistic prediction initially of all ancient applied mathematics models were developed like moving average, exponential smoothing, and autoregressive integrated moving average. Among all, ARIMA became well liked for linear forecasting [1, 10–13].

In this strategy, the given measuring information is introductory checked for stationarity and if don't seem to be then the differencing operation is performed. On the off likelihood that the data are still nonstationary, differencing is a new performed until the information are eventually created stationary [1, 10]. On the other hand differencing is carried out d times, the combination request of the ARIMA philosophy said to be d. The resultant output is sculptural as associate autoregressive moving average (ARIMA) measuring as takes after [1, 10–12, 14, 15]. The information value at any denoted time t, say y_t, is taken under consideration as a function of the earlier p information values, say $y_{t-1}, y_{t-2}, \ldots, y_{t-p}$ and therefore the errors at times $t, t-1, \ldots, t-q$ say $n_t, n_{t-1}, \ldots, n_{t-q}$. The generated ARMA equation is shown in Eq. 1 [1, 2, 10, 12, 13, 16, 17]. In Eq. 1, a_1 to a_p be the autoregressive (AR) coefficients and b to b_q comprises the MA coefficients. So the statistic model is denoted as ARIMA (p, d, q) because it is a mixture of AR and MA. The ARMA model accept that the blunder arrangement n_t is common and that is Gaussian circulated, the difference of mistake is furthermore a model parameter [1, 10] The ARIMA displaying technique consists of three stages: (a) recognizing the model request, i.e., recognizing p and q; (b) assessing the model coefficients; (c) forecast the data [1, 10, 12–14, 16].

The model coefficients are assessed utilizing the Box–Jenkins strategy. At long last, the model coefficients are measurable and the future estimations of the measurement are anticipated utilizing the accessible past data values and hence the model coefficients. ARIMA models anticipate straight measurement data with wonderful accuracy [1, 10–14, 17–20].

$$y_t = a_1 y_{t1} + a_2 y_{t2} + \cdots + a_p y_{tp} + n_t + b_1 n_{t1} + \cdots + b_q n_{tq} \tag{1}$$

In the autoregressive integrated moving average model, the long-term value of a variable is thought to be a linear operator of numerous past observations and random errors [1, 2, 10, 12, 17, 18, 20–22]. ARIMA consolidates an assumption that the real information is regularly straight this suspicion is a reward and disadvantage each for ARIMA because it offers smart forecast result for linear statistic and it cannot deal with nonlinear data [1, 2, 10–12, 17, 18, 20, 21, 23–25].

3 Soft Computing Approach Models

Soft computing is becoming extremely popular and it is getting used in all the fields for obtaining sensible results. And as from literature survey, it proves that ancient models do not seem to be capable to handle nonlinear statistic knowledge properly. Thus, soft computing and intelligence system are a smart alternative for statistic forecasting [26–29].

3.1 Support Vector Machine (SVM) Model

Soft computing approach is Support Vector Machine here. After the introduction of an alternative called loss function, we have use SVM for time series forecasting also [6, 8, 20, 30–34]. The advantage of SVM to resolve nonlinear data problem proves it capable to use for statistic forecasting [32, 33]. The basic plan of SVM as shown in Fig. 1 plots the data x to high-dimensional feature space F by nonlinear plotting space so do linear regression [6, 8, 20, 30–38]. Suppose we have given set of training that consists of n data points $G = \{x_i, d_i\}_{i=1}^{n}$ [34] along with input data $x_i \epsilon R^p$, p consists of a total number of patterns of data and data generated and $d_i \epsilon R^p$ is the result [34].

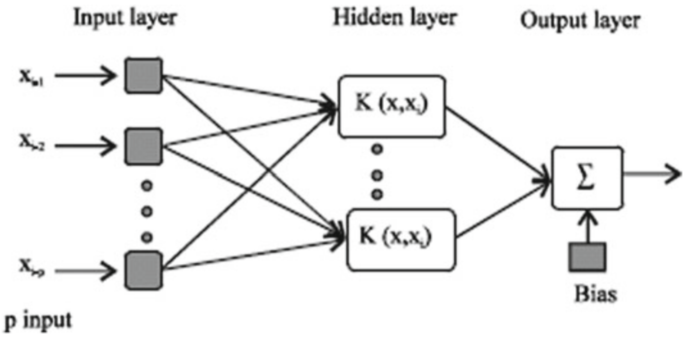


Fig. 1 The architecture of SVM (*Source* [33, p. 952])

The SVM regression as shown in the papers [6, 8, 30, 32, 34, 36] can be calculated approximately by the below function Eq. 2

$$f(x) = w\phi(x) + b, \quad \phi : R^n \to F, w \in F \text{ [32, 36]} \tag{2}$$

where b denotes scalar threshold [36]; ϕ denotes the high-dimensional feature house that is nonlinearly mapped from the input space x. Thus, the regression inside the high-dimensional feature house corresponds to nonlinear regression in low dimension input house, which ignores the real number computation among w and ϕ in the high-dimensional feature house. The coefficients w and b calculable by minimizing the regularized perform as shown in Eq. 3 [8, 20, 31, 32, 36].

$$R(C) = \frac{1}{2}||w||^2 + \frac{C}{n}\sum_{i=1}^{n} L_\varepsilon(d_i, y_i) \tag{3}$$

where

$$L_\varepsilon(d_i, y_i) = \begin{cases} |d_i - y_i| - \varepsilon, & \text{if } |d_i - y_i| \ge \varepsilon \\ 0, & \text{otherwise} \end{cases} \tag{4}$$

To find out the assessment of w and b Eq. 3 is converted to the primal function generated by Eq. 5 by introducing the new positive slack variable ξ and ξ^* as given [5, 6, 34, 36]:

$$\text{Mininmize } R\left(w, \xi^*\right) = \frac{1}{2}||w||^2 + \frac{C}{n}\sum_{i=1}^{n}\left(\xi_i + \xi_i^*\right) \tag{5}$$

$$\text{subject to}\begin{cases} d_i - w\phi(x_i) - b_i \le \varepsilon + \xi_i \\ w\phi(x_i) + b_i - y_i \le \varepsilon + \xi_i^* \\ \xi_i \xi_i^* \ge 0 \end{cases} [39]$$

The term $\frac{1}{2}||w||^2$ denotes the weights vector norm, d_i denotes the required output and C defined as regularized constant [33, 34, 36]. Finally, by introducing Lagrange multipliers as given in [30] and expanding the optimality constraints, the choice operate generated by the equivalent weight of Eq. 2 denotes the subsequent express form [40]

$$f\left(x, \alpha_i, \alpha_i^*\right) = \sum_{i=1}^{n} \left(\alpha_i - \alpha_i^*\right) K(x, x_i) + b \, [39] \tag{6}$$

In Eq. 6, α_i and α_i^* are called Lagrange multipliers [39]. They fulfill the $\alpha_i \times \alpha_i^* = 0$ and $\alpha_i \ge 0$ and $\alpha_i^* \ge 0$ where $i = 1, 2, \ldots, n$ [30, 35, 40] and are generated by maximizing the dual function of Eq. 7, and the maximal dual function generated in Eq. 7 that denotes the below form [36]:

$$R\left(\alpha_i, \alpha_i^*\right) = \sum_{i=1}^{n} d_i\left(\alpha_i - \alpha_i^*\right) - \varepsilon \sum_{i=1}^{n} \left(\alpha_i + \alpha_i^*\right) - \sum_{i=1}^{n} \cdot \sum_{j=1}^{n} \left(\alpha_i - \alpha_i^*\right)\left(\alpha_j - \alpha_j^*\right) K\left(x_i, x_j\right) \tag{7}$$

With constraints, [36]

$$\sum_{i=1}^{n} \left(\alpha_i - \alpha_i^*\right) = 0, \begin{matrix} 0 \le \alpha_i \le C & i = 1, 2, \ldots, n \\ 0 \le \alpha_i^* \le C & i = 1, 2, \ldots, n \end{matrix} \, [36]$$

$K\left(x_i, x_j\right)$ denotes kernel function [39]. The output of the kernel function is equal to join multiplication of two vectors x_i and x_j inside the feature house $\phi(x_i)$ and $\phi(x_j)$, i.e., $K\left(x_i, x_j\right) = \phi(x_i)$ times $\phi(x_j)$ [39–41]. Generally, Gaussian function denotes kernel function and the equation is shown as below [32, 36]

$$K\left(x_i, x_j\right) = -\exp\left(\frac{||\left(x_i - x_j\right)||^2}{2\sigma^2}\right) [36, 39] \tag{8}$$

4 Proposed Work

In this paper, a fresh novel hybrid model has been planned. The system includes ARIMA and SVM. Auto-Regressive Integrated Moving Average (ARIMA) is con-

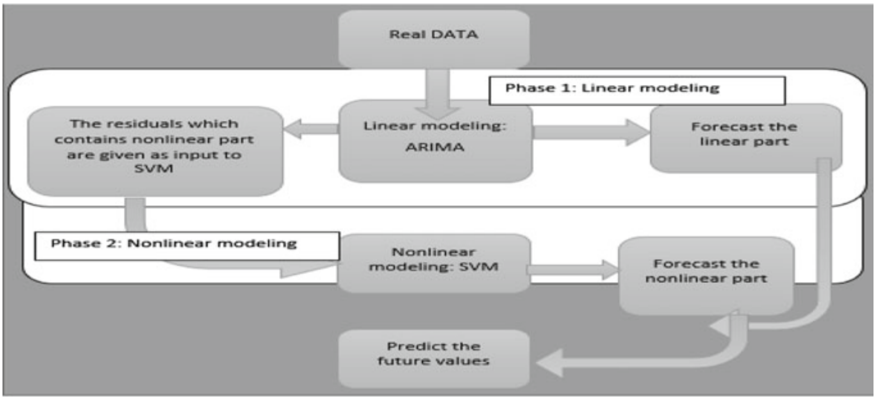


Fig. 2 Proposed model

sidered as one of the popular linear models in statistic prediction throughout the beyond three decades [21, 22]. Current analysis work in prediction along with Support Vector Machine (SVM) suggests that SVM are usually a promising completely different to the traditional linear ways. ARIMA models and SVM together typically compared along with mixed conclusions on the basis of the prevalence for predicting performance [21, 22]. In this paper, a hybrid approach that mixes each ARIMA and SVM model is planned to find out the benefit of the distinctive power of ARIMA and SVM models in linear and nonlinear modeling. Many ARIMA and ANN hybrid models have been planned early which they have given higher results but the accuracy of the model remains a research gap and research tells that SVM could also be a better completely different to ANN so we tend to use SVM in our system. The planned model diagram is shown below in Fig. 2. The proposed model is divided into two phase:

1. Linear Modeling
2. Nonlinear Modeling

Linear modeling

Here, real data is made up of both linear and nonlinear part is taken as input as shown in Eq. 9:

$$Y_{t=L_t+N_t} \tag{9}$$

where Y_t is real data, L_t denotes linear part, and N_t denotes the nonlinear part of data [42]. The assumption is made that real data is a combination of linear and nonlinear data as shown in Eq. 10. Then **Step 1:** First ARIMA is used to design the linear part of data and it produces two outputs: linear forecast values and the residuals.

$$L_t = \left[\sum_{i=1}^{p} \varphi_i Z_{t-i} - \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} \right] + \varepsilon_t + e_t = L_t^{\wedge} + e_t \tag{10}$$

where $L_t^{\wedge}$ is the linear part is forecast value and e_t is the residual part [36] which can be calculated using Eqs. 11 and 13:

$$e_t = Y_t - L_t^{\wedge} \tag{11}$$

Step 2: This residual contains the nonlinear data is given as input to SVM for nonlinear modeling.

Nonlinear modeling

Step 3: Now SVM is used for nonlinear modeling as is best for nonlinear modeling. The residuals from ARIMA are given as input to SVM and it produces the output as forecast values for the nonlinear part by the Eq. 12 as shown below.

$$N_t^{\wedge} = f(e_{t-1}, \ldots, e_{t-n}) + \varepsilon_t \tag{12}$$

Step 4: Now to predict the future the forecast values from ARIMA and SVM are combined and given as input as shown in Eq. 13 and it predicts future values.

$$Y_t^{\wedge} = L_t^{\wedge} + N_t^{\wedge} \tag{13}$$

Step 5: After that, the accuracy of the model is checked by performances measures like MSE.

5 Dataset and Result Analysis

Here, the derived model has been checked for different datasets and the outputs are compared with other models outputs too. The proposed model work is done using R programming.

5.1 Sunspot Data

The sunspot data [3, 10, 11, 21] have been used that consists of the yearly number of sunspots detected on the face of sun ranging from 1700 to 1987, consisting of a total of 288 observations. The plot of particular time series is shown in Fig. 3 says that there is a cyclical pattern that denotes a mean cycle of about 11 years [21, 22].

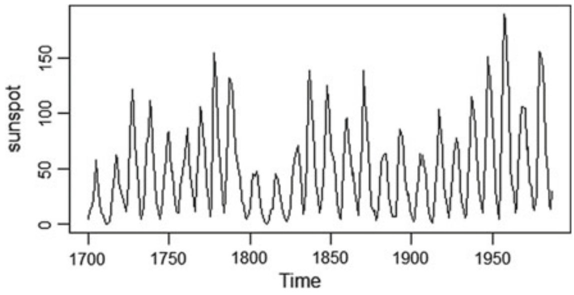


Fig. 3 Sunspot data R plot

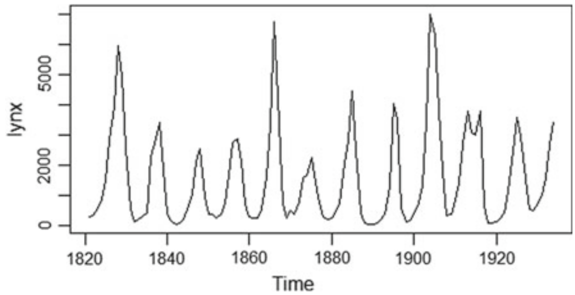


Fig. 4 Lynx data R plot

5.2 Canadian Lynx Dataset

The lynx series [10, 11, 13, 18, 22] consists of the total number of lynx collected per year in the Mackenzie River district of Northern Canada. The lynx data are plotted in Fig. 4 that denotes time period approximately up to 10 years. The datasets consist of total 114 observations, referring to the period of 1821–1934 [21, 22].

5.3 US Monthly Electricity Dataset

The US monthly electricity series [11, 17] contains monthly electricity net generation measured in billions of kilowatt hours (kWh) from January 1973 to October 2010. The dataset has total 454 observations. The plot of series is shown in Fig. 5.

And from the graphs Figs. 6, 7 and 8. And tables of different model comparison for 3 datasets, i.e., Tables 1, 2 and 3 we can see that the proposed model has shown promising accuracy over past models.

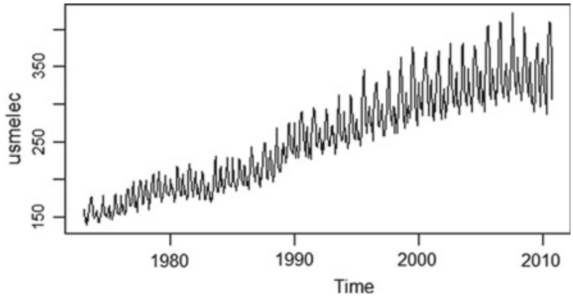


Fig. 5 US monthly electricity data R plot

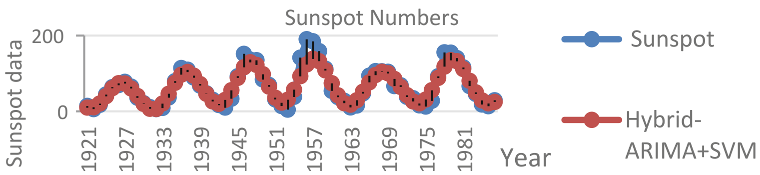


Fig. 6 Sunspot data proposed model R graph

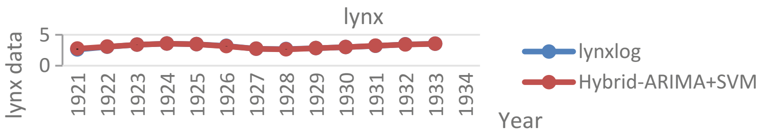


Fig. 7 Lynx data proposed model R graph

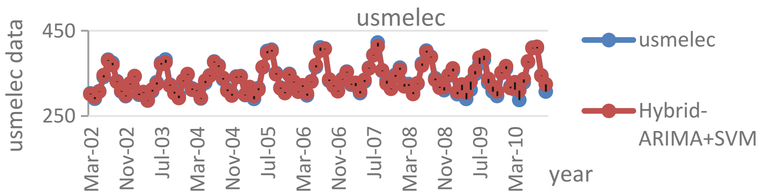


Fig. 8 US electricity data proposed model R graph

6 Conclusion and Future Work

And the results are outstanding for novel hybrid proposed model compared to previously proposed models. The proposed model works well with all type of time series datasets and also gives more accurate results as we can see in tables and graphs above. Thus, SVM can be used as a good alternate of various other models for a hybrid model to get best time series results. But still, SVM parameter selection is an open research gap and needs some good work for better selection process as if

Table 1 Sunspot data result in comparison

Sunspot [221:67] = 288		
Model	MSE	MAE
ARIMA	253.87	17.66
ANN	315.79	12.26
ARIMA ANN [1]	290.19	12.78
ARIMA ANN [2]	240.20	12.10
SVM	285.91	12.00
ARIMA SVM	206.52	11.66

Table 2 Lynx data result in comparison

Lynx [100:14] = 114		
Model	MSE	MAE
ARIMA	0.025	0.123
ANN	0.034	0.148
ARIMA ANN [1]	0.021	0.115
ARIMA ANN [2]	0.014	0.096
SVM	0.076	0.070
ARIMA SVM	0.012	0.019

Table 3 US electricity data result comparison

Usmelec [350:104] = 454		
Model	MSE	MAE
ARIMA	44.80	4.939
ANN	49.74	5.427
ARIMA ANN [1]	78.95	6.696
ARIMA ANN [2]	79.98	7.654
SVM	119.70	6.451
ARIMA SVM	**36.63**	**4.376**

parameters are not selected properly SVM can lead to poor results. Further in the future, we can try other soft computing methods like ANFIS to achieve still more accurate results.

References

1. Babu CN, Reddy BE (2014) A moving-average filter based hybrid ARIMA–ANN model for forecasting time series data. Appl Soft Comput 23:27–38
2. Li C, Chiang T-W (2013) Complex neurofuzzy ARIMA forecasting—a new approach using complex fuzzy sets. Fuzzy Syst IEEE Trans 21:567–584
3. Zounemat-Kermani M, Teshnehlab M (2008) Using adaptive neuro-fuzzy inference system for hydrological time series prediction. Appl Soft Comput J 8:928–936

4. Khashei M, Bijari M (2010) An artificial neural network (p, d, q) model for timeseries forecasting. Expert Syst Appl 37:479–489
5. Chen K-Y, Wang C-H (2007) A hybrid SARIMA and support vector machines in forecasting the production values of the machinery industry in Taiwan. Expert Syst Appl 32:254–264
6. Sapankevych N, Sankar R (2009) Time series prediction using support vector machines: a survey. IEEE Comput Intell Mag 4:24–38
7. Sanghani A, Bhatt N, Chauhan NC (2016) A review of soft computing techniques for time series forecasting. Indian J Sci Technol 9
8. Agrawal RKRA (2013) An introductory study on time series modeling and forecasting. CoRR 1302.6613
9. CiteULike Everyone's library. The underlying motivation for exploiting SVM is that the ability of this technique to accurately forecast statistic knowledge once the underlying system processes are usually nonlinear, non-stationary and not outlined a-priori. Retrieved from http://www.citeulike.org/
10. Babu CN, Reddy BE (2014) A moving-average filter based hybrid ARIMA–ANN model for forecasting time series data. Appl Soft Comput 23:27–38
11. Wang L, Zou H, Su J, Li L, Chaudhry S (2013) An ARIMA-ANN hybrid model for time series forecasting. Syst Res Behav Sci 30:244–259
12. Khashei M, Bijari M (2011) A novel hybridization of artificial neural networks and ARIMA models for time series forecasting. Appl Soft Comput 11:2664–2675
13. Zhang GP (2003) Time series forecasting using a hybrid ARIMA and neural network model. Neurocomputing 50:159–175
14. Li CS, Chiang TW (2013) Complex neurofuzzy ARIMA forecasting—a new approach using complex fuzzy sets. IEEE Trans Fuzzy Syst 21:567–584
15. Sakurai Y, Faloutsos C (2015) Mining and forecasting of big time-series data. 919–922
16. Sciences RB, Sato RC (2013) Disease management with ARIMA model in time series Gerenciamento de doenças utilizando séries temporais com o modelo. ARIMA 11:128–131
17. Khashei M, Bijari M (2010) An artificial neural network (p, d, q) model for timeseries forecasting. Expert Syst Appl 37:479–489
18. Khashei M, Bijari M (2011) A novel hybridization of artificial neural networks and ARIMA models for time series forecasting. Appl Soft Comput J 11:2664–2675
19. De Gooijer JG, Hyndman RJ (2006) 25 years of time series forecasting. Int J Forecast 22:443–473
20. Goel N, Singh S, Aseri TC, Goel N, Singh S, Aseri TC (2013) A review of soft computing techniques for gene prediction. ISRN Genomics 2013:1–8
21. Zhang GP (2003) Time series forecasting using a hybrid ARIMA and neural network model. Neurocomputing 50:159–175
22. Managed DNS | Dyn. Retrieved from https://dyn.com/dns/
23. Khashei M, Bijari M (2012) A new class of hybrid models for time series forecasting. Expert Syst Appl 39:4344–4357
24. Peng H-W, Wu S-F, Wei C-C, Lee S-J (2015) Time series forecasting with a neuro-fuzzy modeling scheme. Appl Soft Comput 32:481–493
25. Wen X, Academy C (2015) A comparative study of artificial neural network, adaptive neuro fuzzy
26. Mandal SN, Choudhury JP, Chaudhuri SRB, De D (2008) Soft computing approach in prediction of a time series data. J Theor Appl Inf Technol 8:1131–1141
27. Bautu E, Barbulescu A (2013) Forecasting meteorological time series using soft computing methods: an empirical study. Appl Math 7:1297–1306
28. Meryem O, Ismail J, Mohammed E-M (2014) A comparative study of predictive algorithms for time series forecasting. In: 2014 third ieee international colloquium in information science and technology (CIST), pp 68–73
29. Rojas I, Palomares H (2004) Soft-computing techniques for time series forecasting. In: Proceedings of the European symposium on artificial neural networks, pp 93–102

30. Xiang L, Zhu Y, Tang G (2009) A hybrid support vector regression for time series forecasting. In: WRI world congress on software engineering, WCSE'09, pp 161–165
31. Gunn SR (1998) Support vector machines for classification and regression by
32. Xiang L, Tang G, Zhang C (2008) Simulation of time series prediction based on hybrid support vector regression. In: Fourth international conference on natural computation, 2008. ICNC'08, pp 167–171
33. Samsudin R (2010) A comparison of time series forecasting using support vector machine and artificial neural network model. J Appl Sci 10:950–958
34. Welcome to Science Alert. Retrieved from http://scialert.net/
35. Chen K-Y (2007) Forecasting systems reliability based on support vector regression with genetic algorithms. Reliab Eng Syst Saf 92:423–432
36. Xiang LXL, Zhu YZY, Tang GTG A hybrid support vector regression for time series forecasting. In: 2009 WRI world congress on software engineering, vol 4, pp 161–165
37. Gunn SR et al (1998) Support vector machines for classification and regression. ISIS technical report, vol 14
38. Vijayalaksmi DP, Babu KSJ (2015) Water supply system demand forecasting using adaptive neuro-fuzzy inference system. Aquat Procedia 4:950–956
39. Bao Y, Liu Z (2006) A fast grid search method in support vector regression forecasting time series. In: Corchado E, Yin H, Botti V, Fyfe C (eds) Intelligent data engineering and automated learning—IDEAL 2006: 7th international conference, Burgos, Spain, 20–23 Sept 2006. Proceedings. Springer, Berlin, Heidelberg, pp 504–511
40. Li Y, Fang T (2003) Rough set methods for constructing support vector machines. In: Wang G, Liu Q, Yao Y, Skowron A (eds) Rough sets, fuzzy sets, data mining, and granular computing: 9th international conference, RSFDGrC 2003, Chongqing, China, 26–29 May 2003 Proceedings. Springer, Berlin, Heidelberg, pp 334–338
41. Zhang X, Zhang T, Young AA, Li X (2014) Applications and comparisons of four time series models in epidemiological surveillance Data. PLoS ONE 9:e88075
42. Chen KY, Wang CH (2007) A hybrid SARIMA and support vector machines in forecasting the production values of the machinery industry in Taiwan. Expert Syst Appl 32:254–264

Medical Image Classification Using MRI: An Investigation

R. Merjulah and J. Chandra

Abstract The main objective of the paper is to review the performance of various machine learning classification technique currently used for magnetic resonance imaging. The prerequisite for the best classification technique is the main drive for the paper. In magnetic resonance imaging, detection of various diseases might be simple but the physicians need quantification for further treatment. So, the machine learning along with digital image processing aids for the diagnosis of the diseases and synergizes between the computer and the radiologist. The review of machine learning classification based on the support vector machine, discrete wavelet transform, artificial neural network, and principal component analysis reveals that discrete wavelet transform combined with other highly used method like PCA, ANN, etc., will bring high accuracy rate of 100%. The hybrid technique provides the second opinion to the radiologist on taking the decision.

1 Introduction

Over the earlier several decayed, ailments have dropped before the scythe of the intelligence of the human in the system of the biomedical developments with the understanding of the various diseases. A qualified radiologist visually examines the medical images and further recognizes the signs of the various diseases [1]. Image analysis supports the semiautomatic or automatic technique for the illustration of the acquired medical images. Since the generation of the clinical data is more, it is difficult to physically describe and classify data in suitable time. Despite the specialist's skills and experience, the qualitative manual analysis is based on the vision of the human system. The incapability of the human eye cannot able to differentiate

R. Merjulah (✉) · J. Chandra
Christ University, Bangalore, India
e-mail: r.merjulah@res.christuniversity.in

J. Chandra
e-mail: Chandra.j@christuniversity.in

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_108

the several tens of levels in the MRI. The inability leads in using the computer-aided machine as the second opinion for the in-depth and high-resolution of MRI images.

Machine learning (MI) is the procedure of image processing, natural language processing and computer vision are the key artificial intelligence (AI) technologies forming the pillars for the radiologists. Current AI an has accurate rate that surpasses the radiologists in some area of narrow study [2]. The strength of the AI is to make the clinical judgment through the accurate detection and classification of the MRI images. MI is for the automatic detection as well as the classification of various diagnosis using MRI. Combining the radiologist and the AI as the hybrid model will lead to promising accuracy in the diagnosis. Incorporation of the MI will become the driving force for the health care in the future and also increases the value of the Radiology in the medical field [3].

Classification algorithm is classified into supervised and unsupervised methods. To train the dataset in the supervised method, the available labels for every data points are used for the classifier. In unsupervised methods, the datasets are trained using the restricted prior information of the data. The unsupervised method which analyzes the arrangement of the data to infer labels. In the medical imaging, the label often initiates from the spatially determined histological examinations. In a classical classifier, every object is trained and tested by the vector of the features and a rule is applied to classify the tested vector. In the image classification problem, these feature vectors are obtained from the images using every pixel that corresponds to one feature. Although there are many classification techniques for the medical images, most of them are not suitable for various medical images. Finding out the best method for the medical image dataset is the challenging task.

The classification process work flows like image acquisition, preprocessing, feature extraction, classification, and evaluation. Image acquisition is nothing but the selection of images that ranges as either CT or MRI or X-rays, etc. Preprocessing is the process of executing the raw images for the best performance of the dataset as images. Preprocessing phase includes cropping, filtering, noise removal, etc. Feature extraction involves with the feature estimation and the selection of the features. The literature provides some features like gabor, texture, wavelet histogram, etc. Every aspect explains the content of the images. Feature extraction is to extract the prominent features that correspond to the different classes of objects. Therefore improved feature extraction brings high accuracy in the classification technique, where the classification is divided into three division namely neural network, texture classification, and data mining technique.

In general, the complete diagnosis process includes different stages such as preprocessing, segmentation, feature selection, feature extraction, classification, etc. The purpose of the medical image analysis is to automate the process of measuring and for meaningful interpretation. The main objective of the work is to identify the best classification technique for MRI on various diagnosis.

2 Related Works

Recently, a lot of researches based on the classification technique include either the direct implementation of the best image classification technique or some other research coming out with specific problem-solving technique using hybrid model. Since, the image processing for the medical images like computer tomography (CT), electrocardiogram (ECG), MRI, etc., is totally more challenging that the other image processing.

Kumar et al. [4] have proposed a hybrid model on the brain MRI tumor images which contains discrete wavelet transform (DWT) for feature extraction, genetic algorithm for decreasing the number of features and used support vector machine (SVM) for the brain tumor classification. The accuracy varies from 80 to 90% for 25 dataset of MRI brain images.

Zhang et al. [5] have proposed a hybrid model for the classification of MRI brain images as normal or abnormal. The feature extraction includes a DWT from images, the reduction of the features includes the principle component analysis (PCA), and the reduced features are passed through the forward neural network (FNN). An improved artificial bee colony (ABC) algorithm named scaled chaotic artificial bee colony (SCABC) is used for the optimization of the parameters. The experiment is tested with 66 brain images and achieved 100% accuracy on MSE for the classification.

Chaplot et al. [6] have proposed wavelets to neural network self-organized maps (SOM) and SVM for the classification of normal and abnormal brain MRI. The classification achieves 94% on SOM and 98% on SVM for 52 MRI brain image dataset.

Saritha et al. [7] have proposed an integrating wavelet entropy based spider web plots combined with the probabilistic neural network for the classification of the brain MRI. The classification yields 100% accuracy on the brain dataset.

Gupta et al. [8] have proposed DWT for the feature extraction and PCA as feature extraction and used various classification technique like SVM, K-nearest neighbor (KNN), classification and regression tree, and random forest (RA). The SVM without the PCA yields 88% accuracy on the classification of cerebral Tumor.

Duchesne et al. [9] have proposed SVM for the automatic classification of neurodegenerative diseases, specifically Alzheimer's dementia (AD). The classification is based on the separation of 75 probable AD and 75 age-matched normal controls. The result of classification achieves 92% based on the least squares optimization.

Sayed et al. [10] have proposed KNN and linear discriminant analysis (LDA) for the classification of two groups namely benign and malignant tumors. The examined classification of the breast tumors with the two methods reveals the significant accuracy compared with the pathological analysis and resubstitution error.

Gatidis et al. [11] have proposed a classification of the multidimensional medical images for the local prostate cancer. To train the SVM, a single dataset was applied to the spatial constrain fuzzy C-Means (SFCM) algorithm. The proposed model is verified with the false positive, false negative and accuracy where determined and compared with the manual tumor delineation. The hybrid model of the combination of SFCM and SVM yields a better result than the unsupervised SFCM alone.

Garro et al. [12] have proposed ANN with ABC for the classification of the DNA microarray. ABC is used to train the ANN with the reduced genes. ABC is used mainly for the dimensionality reduction for selecting the best set of genes.

Dandil et al. [13] have proposed a spatial fuzzy C-means technique for the segmentation and SVM for the classification of the brain tumor into malign and benign which yields the accuracy of 91.49%.

Dahshan et al. [14] have proposed feedback pulse-coupled neural network for the image segmentation and for the identification of the ROI. The further work on the feature extraction is done using the DWT. Further, for the reduction of the dimensionality features, the PCA technique is used and the reduced features are passed through the back-propagation neural network for the classification of abnormal or normal brain based on the selection of the feature parameters.

Kalbkhani et al. [15] have proposed a multi-cluster feature selection technique for the feature selection on the brain MRI. The features are obtained from the DWT. Further, the features are applied to K-nearest neighbor (KNN) for the classification of normal MRI or one of seven various illness.

Amien et al. [16] have proposed an automated brain tumor diagnosis for MRI. First step involves the reduction of the noise and increasing the contrast of the images. Second step involves the extraction of the texture features and for the dimensionality reduction, the PCA is used. Finally, the reduced features are passed through the back-propagation neural network (BPNN) for the classification of brain images as edema or normal or cancer and the result achieves the accuracy of 96.8%.

Through the above literature review, it is observed that the hybrid machine learning models bring out high accuracy for the successful classification.

3 Methodology

In the section, the algorithms for the analysis of the medical MRI images are explained based on the classification techniques like SVM, DWT, ANN, and PCA. The study is to investigate the best classification technique for the MRI and determines that DWT with other highly used algorithms like ANN and PCA yields high accuracy rate.

3.1 Support Vector Machine

SVM is the supervised machine learning algorithm which is used for the regression as well as for the classification. In the SVM classification, the plotting happens with every data item as point in the n dimensions where the number of features is n. Then the classification is performed using the hyperplane that differentiates the classes. Manually selecting the hyper plane is the challenging task. Choose the hyperplane that segregates the classes better. In SVM for the automatic hyperplane selection, a technique called Kernel trick is used. Kernel trick transforms the low dimension input space to the high dimension space. It is commonly used in the nonlinear separation problem.

Mathematical calculation falls into the following steps

Step 1: Dataset D is inserted into the SVM for the classification

The data is composed of n vectors x_i. The x_i value is associated with the y_i that indicates the elements which belong to the +1 class or −1 class. y_i belong to either −1 or +1.

$$\text{Dataset } D = \left((x_i, y_i) | x_i \in R_p, y_i \in \{-1, 1\}\right)_{i=1}^{n} \tag{1}$$

Step 2: Selecting the hyperplane with no points

The hyperplanes are defined as H such that

$$H_1 \text{ defines as } w.x_i + b = +1 \tag{2}$$

$$H_2 \text{ defines as } w.x_i + b = -1 \tag{3}$$

$$H_0 \text{ defines as } w.x_i + b = 0 \tag{4}$$

where H_1 and H_2 are the classification planes and H_0 is the in-between median. The weight vector is w, input vector is x, and bias is b.

$d+$ mentioned as the short distance to the close positive point.
$d-$ mentioned as the short distance to the close negative point.

3.2 *Discrete Wavelet Transform*

Wavelets are to transform the signal hooked on to the set of basic functions. The DWT is to transform a discrete time signal to discrete wavelet. DWT is the highly efficient and decomposition of signals. DWT is the multi-resolution exploration and it decomposes the images in scaling function and wavelet coefficient. Wavelet converts the images into series of wavelets which stores more efficiently than the pixel blocks. Digital filtering technique for the digital signal for DWT timescale representation. The signal is analyzed which is passed through the filters with cutoff frequency for various scales that reduce the time taken during the computation process [17]. In two dimensional DWT, the rows are first processed with one level of decomposition. The decomposition separates the first half that stores the details of the coefficients and the other half stores the average coefficients. The procedure is continued for the column results in sub-bands.

3.2.1 Haar Transform

Haar transform is a modest category of wavelet transform. In discrete, the Haar wavelets are linked to the scientific operation named as haar transform. This transform provides the prototype for the other wavelet transform. Similarly, other transforms decompose the discrete signal into two sub-signals of the length. The running average is one signal and the other is the fluctuation or running difference.

3.2.2 Daubechies Transform

The Daubechies transform defines the similar way as haar transform by processing the running average and the differences between the scalable product with the scaling signals and the wavelets [18]. The wavelet balances the occurrence responses but it is nonlinear phase response. Daubechies transform uses overlapping windows, so, high-frequency coefficient spectrum that reflects the high-frequency changes.

Basically, the medical images need the accuracy without the loss of the information. DWT is based on the discrete time signal that provides the multi-resolution without loss of information on the MRI data.

3.3 Artificial Neural Network

ANN is based on the supervised family. ANN acquires the knowledge through the connection of the network. The general steps of the neural network are to define the fixed number of nodes in the input, output, and in the hidden layers. The following steps are used for training the ANN:

Step 1: Push the input for the processing.
Step 2: Every input has to be weighted, which is nothing but multiplied by random value of either +1 or -1.
Step 3: Add all the weighted input values.
Step 4: Pass the sum through activation function for producing the output of the network.

ANN has the learning ability for various classification problem with back-propagation [19], and improved back-propagation algorithm [20].

3.4 Principal Component Analysis

PCA is the usual statistical model used for the reduction of the dimensionality of the data. There are many algorithms for the reduction of the features in the images and choosing important ones. The PCA is the best among that for the reduction of the features and improves the accuracy result in the classification task.

If there are n is the observation and p is the variable, then the number of PCA is

$$\min(n-1, p) \tag{5}$$

PCA preserving as many variables that has been translated into the new variable that are the linear functions on the original dataset which increases the variance that are not overlapped with each other. The new variable reduces the eigenvector problem. The linear combination of matrix in

$$\sum_{j=1}^{p} a_j x_j = xa \tag{6}$$

where p is the data value, n is the dimension vectors of $x_1, x_2 \ldots, x_p$, $n \times p$ is the data matrix x, whose jth column is the x_j vector of the jth variable, constant vector is $a_1, a_2, a_3, \ldots, a_p$.

The steps in the PCA includes,

Step 1: Using the dataset "S" calculate the mean value
Step 2: Subtract the mean using "S". A new matrix "A" value is obtained.
Step 3: Process the covariance "c" from "A" using $C = AA^{\mathrm{T}}$ where the computation process through

$$1, \ldots, l,\ A_k \in R,^N \sum_{k=1}^{l} A_k = 0,$$

$$\text{Covariance matrix is } c = \frac{1}{l} \sum_{i=1}^{l} A_i A_i^{\mathrm{T}} \tag{7}$$

Step 4: The values of the eigen will be obtained through the covariance matrix where $[V_1, V_2, \ldots, V_n]$.
Step 5: Using the covariance matrix C eigenvectors are calculated.
Step 6: Linear combination of eigenvectors of vector S or $S - \bar{S}$

$$S = b_1 u_1 + b_2 u_2 + b_3 u_3 \cdots + b_n u_n \tag{8}$$

Since the covariance form is $[V_1, V_2, \ldots, V_n]$.
Step 7: The largest eigenvalue gives the lower dimensional dataset

$$S = \sum_{i=0}^{1} b_i u_i,\ 1 < N \tag{9}$$

The lower component dimensional space is the principal components. The principal components will be independent in the normal case distribution.

3.5 Self-organizing Map

SOM is developed by Kohonen [21], the SOM texture classification algorithm [22], and the steps involved in SOM are as follows

Step 1: Initialize the three nodes which includes connection weights, output nodes, and input nodes. Where the input vector is N, M is the two-dimensional map as output vector, the weight is w_{ij} from the input nodes to the M output nodes.
Step 2: Place every document in order. Set the value 1 for the document which has the corresponding term. Set value 0 if the document has no corresponding term. The document is placed several times.
Step 3: The distance is computed using the Euclidean distance d_j between the input vector and the output vector.

$$d_j = \sum_{i=0}^{N-1} (x_i(t) - w_{ij}(t))^2 \tag{10}$$

where the value of $x_i(t)$ can be either 0 or 1 that depends on the ith document term at the time t. The vector that represents j map node in the vector document space is w_{ij}. The interpretation of the weight from the input node is i to the output node j in the neutral net.
Step 4: Choose the winning node j and the weights to node j has to be updated and its neighbors to reduce the distance.

$$w_{ij}(t+1) = w_{ij}(t) + \eta(t)\big(x_i(t+1) - w_{ij}(t)\big) \tag{11}$$

where $\eta(t)$ is an error adjusting coefficient $(0 < \eta(t) < 1)$ which decreases the overall time.
Step 5: Label the map regions. After the training of the network, assign the term to every node by selecting the winning term. The similar winning term will be combined to form the group. The resulting map represents the documents that joined with them. The processing time increases because the Steps 2 to 4 are repeated several times.

4 Result Analysis

In the section, the dataset description and the performance evaluation of the DWT, SVM, ANN, KNN, and the hybrid approaches are shown to find the best classification

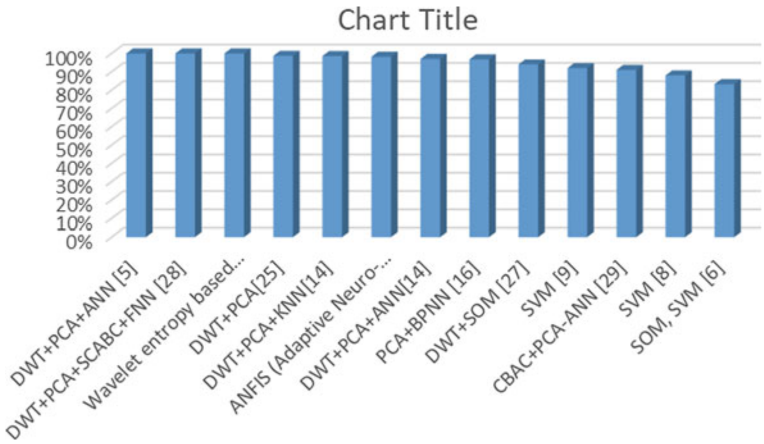


Fig. 1 Performance graph on various classification technique using MRI

model for the classification of the medical MRI data images. The result is compared base on the accuracy.

The whole brain atlas dataset is the information resources of the central nervous system imaging that has the clinical information of the magnetic resonance imaging, X-ray computed tomography, and the nuclear medicine images [23]. The atlas project is part of the radiology department and neurology from Brigham and women's hospital.

Allan Institute is another open source repository that currently collaborated with three entries namely Allen Institute of Cell Science, brain science, and finally the Paul G. Allen Frontiers group. The approaches for the bioscience ranges from the smallest molecular scale to level of the entire system [24].

Accuracy is the probability for the diagnosis testing which can be calculated as follows:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{12}$$

where TP is True Positive for the correct classification of positive cases, TN is True Negative for the correct classification of the negative cases, FP is the False Positives for the incorrect classification of negative cases, and FN is the False Negative for the incorrect classification of the positive cases.

Table 1 shows the performance evaluation of the classification technique and the hybrid approaches with the accuracy. To evaluate the effectiveness based on the accuracy, the machine learning technique has been compared with the existing models in the table to find out the high accuracy classification. Figure 1 shows the performance comparison of the existing system. Finally, after the investigation, it reveals that the DWT along with the highly used machine learning technique like ANN and PCA of hybrid model brings out high accuracy.

Table 1 Performance evaluation of the various classification techniques using MRI

S. No.	Dataset	Approach	Classification accuracy (%)
1	Brain	DWT + PCA + ANN [5]	100
2	Brain	SOM, SVM [6]	83.21
3	Brain	Wavelet entropy based spider web plot + PNN [7]	100
4	Cerebral tumor	SVM [8]	88
5	Alzheimer's dementia	SVM [9]	92
6	Brain	DWT + PCA + ANN [14]	97
7	Brain	DWT + PCA + KNN [14]	98.60
8	Brain	PCA + BPNN [16]	96.8
9	Brain	DWT + PCA [25]	98.75
10	Brain	ANFIS (Adaptive neuro-fuzzy inference system) [26]	98.25
11	Brain	DWT + SOM [27]	94
12	Brain	DWT + PCA + SCABC + FNN [28]	100
13	Brain	CBAC + PCA − ANN [29]	91

5 Conclusion

A study on the recent classification techniques based on the magnetic resonance imaging collectively gets the knowledge and its accuracy is the stepping stone for the researchers for the classification of medical images. The investigation based on the classification of medical images reveals that the discrete wavelet transform along with the highly used classification technique like artificial neural network, principal component analysis, etc., brings out high accuracy of 100%. The survey is designed towards the expansion of image processing and machine intelligence technique. The digital image processing methodologies along with the intelligence technique support the radiologist in effective diagnosis, and as hybrid methods provide the second opinion and the assistance to the radiologist.

References

1. Mohan G, Subashini MM (2018) MRI based medical image analysis: survey on brain tumor grade classification. Biomed Sig Process Control 39:139–161
2. Hinton G (2016) Machine learning and the market for intelligence. In: Machine learning and the market for intelligence conference
3. Dean BL, Drayer BP, Bird CR, Flom RA, Hodak JA, Coons SW, Carey RG (1990) Gliomas: classification with MR imaging. Radiology 174:411–415
4. Kumar S, Dabas C, Godara S (2017) Classification of brain MRI tumor images: a hybrid approach. Procedia Comput Sci 122:510–517. Springer
5. Zhang Y, Wu L, Wang S (2011) Magnetic resonance brain image classification by an improve artificial bee colony algorithm. Prog Electromagn Resolut 116:65–79
6. Chaplot S, Patnaik LM, Jagannathan NR (2006) Classification of magnetic resonance brain images using wavelets as input to support vector machine and neural network. Biomed Sig Process Control 1(1):86–92
7. Saritha M, Paul Joseph K, Mathew AT (2013) Classification of MRI brain images using combined wavelet entropy based spider web plots and probabilistic neural network. Pattern Recogn Lett 34(16):2151–2156
8. Gupta T, Gandhi TK, Gupta RK, Panigrahi BK (2017) Classification of patients with tumor using MR FLAIR images. Pattern Recogn Lett
9. Duchesne S, Caroli A, Geroldi C, Barillot C, Frisoni GB, Collins DL (2008) MRI-based automated computer classification of probable AD versus normal controls. IEEE Trans Med Imaging 27:509–520
10. Sayed AM, Zaghloul E, Nassef TM (2016) Automatic classification of breast tumors using features extracted from magnetic resonance images. Procedia Comput Sci 95:392–398
11. Gatidis S, Scharpf M, Martirosian P, Bezrukov I, Küstner T, Hennenlotter J, Kruck S, Kaufmann S, Schraml C, Fougere C, Schwenzer NF, Schmidt H (2015) Combined unsupervised—supervised classification of multiparametric PET/MRI data: application to prostate cancer. Biomedicine 28(7):26
12. Garro BA, Rodriguez K, Vazquez RA (2016) Classification of DNA microarrays using artificial neural network and ABC algorithm. Appl Soft Comput 38:548–560 Elsevier
13. E. Dandil., M. Cakiroglu., Z. Eksi.,: Computer-aided diagnosis of malign and benign brain tumors on MR images, in ICT Innovations, pp. 157–166, 2014
14. El-Dahshan ESA, Hosney T, Salem ABM (2010) Hybrid intelligent techniques for MRI brain images classification. Digital Sig Process 20:433–44
15. Kalbkhani H, Salimi A, Shayesteh MG (2015) Classification of brain MRI using multi-cluster feature selection and KNN classifier. In: Electrical engineering conference. IEEE, pp 2164–7054
16. Amien MB, Abd-elrehman A, Ibrahim W (2013) An intelligent model for automatic brain-tumor diagnosis based-on MRI images. Int J Comput Appl 72(23):21–24
17. Bute YS, Jasutkar RW (2012) Implementation of discrete wavelet transform processor for image compression. Int J Comput Sci Netw 1:1–5
18. Mahmoud MI, Dessouky MIM, Deyab S, Elfouly FH (2007) Comparison between Haar and Daubechies wavelet transformions on FPGA technology. In: Proceedings of world academy of science, in engineering and technology, vol 20
19. Heermann PD, Khazenie N (1992) Classification of multispectral remote sensing data using a back-propagation neural network. IEEE Trans Geosci Remote Sens 30(1):81–88
20. Nawi NM, Ransing RS, Salleh MNM, Ghazali R, Abdul Hamid N (2010) An improved back propagation neural network algorithm on classification problems. Database Theory Appl Bio-Sci Bio-Technol 118:177–188
21. Kohonen T (1995) Self-organization maps. Springer, Berlin, Heidelberg
22. Raghu PP, Poongodi R, Yegnanarayana B (1995) A combined neural network approach for texture classification. Neural Networks 8(6):975–987

23. Summers D (2003) Harvard whole brain atlas. J Neurol Neurosurg Psychiatry 74(3):288
24. Miller JA et al (2014) BrainSpan Atlas of the developing human brain, in Transcriptional landscape of the prenatal human brain. Nature 508:199–206. https://doi.org/10.1038/nature13185
25. Zhang Y, Wang S, Wu L (2010) A novel method for magnetic resonance brain image classification based on adaptive chaotic PSO. Prog Electromagn Res 109:325–343
26. Roy S, Sadhu S, Bandyopadhyay SK, Bhattacharyya D, Kim T-H (2016) Brain tumor classification using adaptive neuro-fuzzy inference system from MRI. Int J Bio-Sci Bio-Technol 8(3):203–218
27. Chaplot S, Patnaik LM, Jagannathan NR (2006) Classification of magnetic resonance brain images using wavelets as input to support vector machine and neural network. Biomed Sig Process Control 1:86–92
28. Zhang Y, Wu L, Wang S (2012) Magnetic resonance brain image classification by an improved artificial bee colony algorithm. Prog Electromagn Res 130:369–388
29. Sachdeva J, Kumar V, Gupta I, Khandelwal N, Ahuja CK (2013) Segmentation, Feature Extraction and Multiclass brain tumor classification. J Digital Imaging 26(6):1141–1150

Tumor Detection and Analysis Using Improved Fuzzy C-Means Algorithm

R. Swathika, T. Sree Sharmila, M. Janani Bharathi and S. Jacindha

Abstract Formation of abnormal cells in brain serves the major cause of tumor. With estimated deaths of 229,000 as of 2015, it has become an issue to be dealt. The less awareness of brain tumor owes to lots of unaccounted deaths. Thus, we aim in developing an app which could serve the purpose of detecting the tumor and giving additional information related to the detected tumor. This app takes in an MRI image and does preprocessing followed by clustering, segmentation, and binarization. The preprocessing involves the conversion of the image into grayscale and noise filtering. We aim at using improved Fuzzy c-means algorithm for clustering and segmentation. Binarization mainly aims at calculating the tumor size useful for further analysis. The improved fuzzy c-means algorithm overcomes the various constraints of k-means algorithm such as time complexity, processing of noisy images, and memory space.

1 Introduction

Brain tumor segmentation and analysis is a field of interest for many research scholars. But an efficient method for analyzing the tumor is yet to be developed as the existing methods face one or the other drawbacks. As for now, the best segmentation technique is k-means [1] but the efficiency of k-means is limited by various constraints such as processing noisy images, time complexity, and memory space. Fuzzy c-means [FCM] was brought into picture to serve this purpose but neither of them could eradicate these problems in an efficient way. Thus, we propose improved fuzzy c-means algorithm for clustering and segmentation of the image. The IFCM takes the frequency of the pixel values instead of the actual values which had proved to greatly reduce the time complexity and memory space. Awareness about the brain tumor is very less among the people. In a developing nation like our country still a major population is uneducated and they have very little knowledge about tumor. If our system could analyze the tumor more efficiently, we definitely believe that we

R. Swathika (✉) · T. Sree Sharmila · M. Janani Bharathi · S. Jacindha
Department of Information Technology, SSN College of Engineering, Chennai, India
e-mail: swathikar@ssn.edu.in

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_109

could save more lives. Today in this era of technological advancements, we aim in providing efficient and interesting technology which could take MRI image as an input and detect the tumor area and depending upon the size of the tumor detected, it could suggest the stage of the tumor. Localized detection of the tumor could give an output as benign and if it had spread over a number of regions, it could classify it as malignant or secondary. This technology could be used by doctors in their analyzation.

2 Review of Existing Work

Yisu Lu et al. proposed Dirichlet's process mixture model for Automatic Multimodal Brain Tumor Segmentation [2]. The time complexity of the work is $O(n^2)$. Hence, the processing of the high-resolution images could become a tedious and time-consuming job. 1024 by 1024 data matrix could take computational time of 10 M s. It also takes up lots of the memory space.

In automatic image segmentation using convolutional neural network [3], an optimal number of clusters for each unique image is predetermined and hence it enhances the clustered image and serves the best in terms of the accuracy. Our proposed algorithm could be used in combination with this work for the predetermination of number of clusters to be passed an argument to the function performing the IFCM.

Model-based brain and tumor segmentation [4] uses expectation maximization clustering algorithm. This is variant of k-means and hence suffers from the drawbacks of time complexity and memory space. The time complexity is in the order of $O(n * k)$. Thus, here also the time increases in proportion with the input size.

Telrandhe et al. [5] proposed detection of brain tumor from MRI images by using segmentation and SVM. The segmentation and preprocessing of image are done using k-means algorithm. The classification and detection of brain tumor are done using SVM technique. Ghassabeh et al. [6] used IFCM with genetic algorithm optimization for segmentation of brain tissue. Forghani et al. [7] proposed a segmentation of brain tissue using IFCM with particle swarm optimization.

Most of the other applications currently available use either k-means, FCM else hierarchical clustering models for the segmentation followed by the binarization and thus extracting the tumor from the MRI image.

2.1 *Traditional FCM*

Traditional FCM (fuzzy c-means clustering) is a segmentation technique which segments the input image into number of predefined clusters. It is developed by Dunn and changed by Bezdek [8]. Segmentation is done in accordance with the similarity of the pixel values. The advantage of FCM is that it is a soft clustering technique and hence is able to outperform hard clustering technique k-means. The performance evaluation of k-means and clustering algorithms with that of FCM is discussed in detail in Table 1. FCM allows a pixel to belong to two or more clusters and hence provides more accuracy in comparison with other techniques. The following are the steps of traditional fuzzy c-means algorithm.

Algorithm: fcm(x[l..n])
//Clustering by traditional FCM method
//Input: The data matrix x of dimension l by n
//Output: The membership matrix U
Step1: Choose random centroid at least 2 and put values to them randomly.
Step2: Compute membership matrix:

$$U_{ij} = \frac{1}{\sum_{k=1}^{c}\left[\frac{|x_i - c_j|}{x_i - c_k}\right]^{\frac{2}{m-1}}}, \quad \text{where } m > 1, c \text{ cluster's No.}$$

where U is the membership matrix,
C is the centroid matrix.
i, j, k are loop variables.
Step3: calculate the clusters centers:

$$C = \frac{\sum_{i=1}^{n} U^{m} ij * x_i}{\sum_{i=1}^{n} U^{m} ij}$$

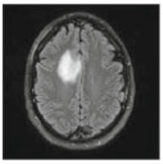
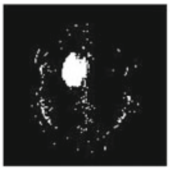
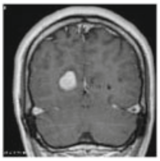
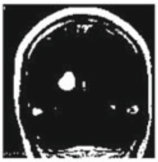
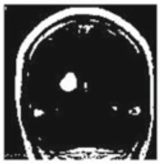
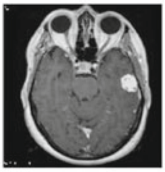
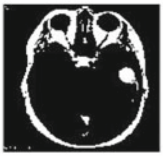
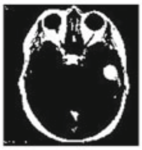
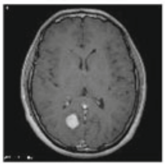
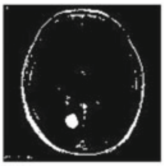
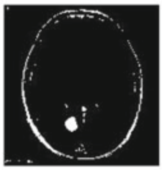
Step4:
if $C^{k-1} - C^{k} < \varepsilon$
then Stop else go to Step2.

This traditional algorithm is an iterative algorithm that suffers from time and memory consumption because it computes membership value for each item in the data.

3 Proposed Work

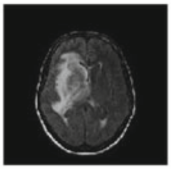
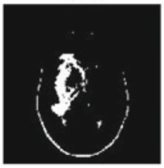
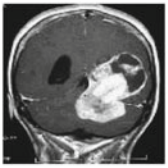
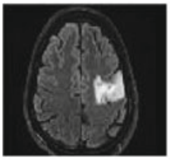
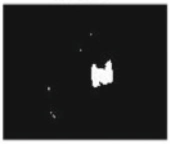
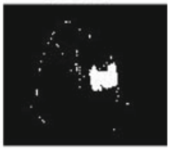
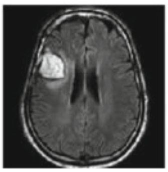
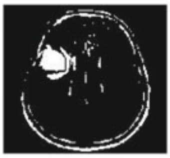
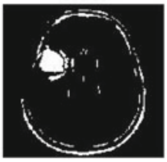
Our app gets the image from the gallery and converts it into gray scale. After filtering, the image is clustered and segmented using IFCM. The IFCM output then undergoes binarization to isolate the tumor with which the size of the tumor is calculated. The steps used for tumor detection are clustering and thresholding. The process diagram is shown in Fig. 1.

Table 1 Comparison of performance evaluation of the proposed clustering algorithm

IMAGES	FCM	K-MEANS	IFCM
			
COMPUTATIONAL TIME	36.415 sec	3.041 sec	2.1764 sec
			
COMPUTATIONAL TIME	71.562 sec	6.692 sec	6.4719 sec
			
COMPUTATIONAL TIME	147.706 sec	10.586 sec	10.414 sec
			
COMPUTATIONAL TIME	290.432 sec	16.55 sec	13.687 sec

(continued)

Table 1 (continued)

IMAGES	FCM	K-MEANS	IFCM
			
COMPUTATIONAL TIME	10.689 sec	2.798 sec	1.728 sec
			
COMPUTIONAL TIME	34.837 sec	18.03 sec	3.153 sec
			
COMPUTATIONAL TIME	20.458 sec	3.065 sec	2.585 sec
			
COMPUTATIONAL TIME	7.824 sec	2.764 sec	2.014 sec

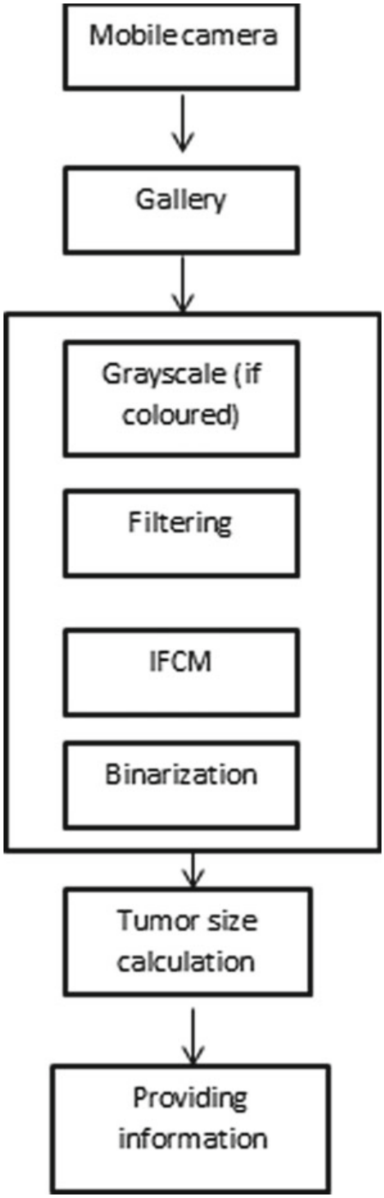


Fig. 1 Process flowchart

3.1 Preprocessing

Converting image into gray valued matrix

The input image in the form of an MRI scanned image is first converted into gray valued matrix. Conversion of image into grey valued matrix can be done by an inbuilt function in MATLAB.

Noise filtering—median filtering

It filters the image and gives a smooth appearance and also prepares it for further processing. Median filtering has proved to be very effective in the field of medical image processing (Figs. 2 and 3).

Fig. 2 Original image

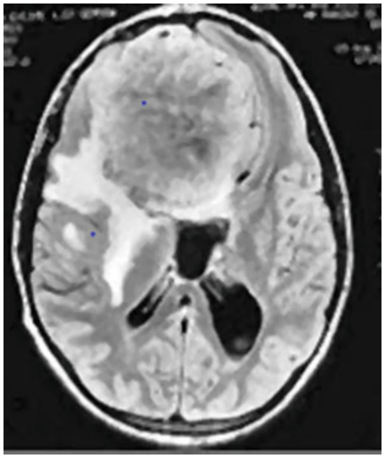


Fig. 3 After median filtering

3.2 Improved Fuzzy C-Means

In the following section, we propose the improved fuzzy c-mean algorithm:

Algorithm: ifcm(data[l..n])
//IFCM method
//Input: data matrix data of dimension l by n
//Output: The membership matrix U
Step1: Let H represent the frequency of each item in Data.
Step2: create vector I=min(Data): max (Data)
Step3: Choose random centroid at least 2.
Step4: Compute membership matrix:

$$U_{ij} = \frac{1}{\sum_{k=1}^{c}\left[\frac{|I_i - c_j|}{I_i - c_k}\right]^{\frac{2}{m-1}}}$$

Where U is the membership matrix,
I is a vector which contains the unique elements in the data matrix I the ascending order from minimum to maximum,
m is a minimum exponential value and is ≥2
c is the centroid matrix
Step5: calculate the cluster center:

$$C = \frac{\sum_{i=1}^{n} U^m * H * 1}{\sum_{i=1}^{n} U^m * H}$$

Step 6: if $C^{(k-1)} - C^k < \varepsilon$ then stop else go to step

In traditional FCM, the membership matric raised to the power m is multiplied with a data points matrix whereas in IFCM [improved fuzzy c-means] frequency matrix H is multiplied. H is N by 1 matrix and N << total no. of. data points and $0 < N < 256$. I is also a N by 1 matrix. Hence on the application of IFCM, the number of iterations for the calculation of new center and new membership matrix decreases rapidly thus reducing the overall time complexity to a great extent. The maximum number of elements to be processed here is 256 whereas in traditional FCM, it could be 256 * 256 = 65,536. The minimum value of m is 2. The proposed algorithm does not depend on whole data of image; it actually depends on data that represent the frequency of each data item in original image's data. A number of frequencies at most are 256.

```
I=unique(data);
H1=[I,histc(data(:),I)];
H=H1(:,2);
tmp1=H.*I;

mf = U.^expo;        % MF matrix after exponential modification
center = mf*tmp1./(sum(mf,2)*ones(1,size(tmp1,1))*H); %new center
dist = distfcm_f(center, data);        % fill the distance matrix
obj_fcn = sum(sum((dist.^2).*mf));  % objective function
tmp = dist.^(-2/(expo-1));      % calculate new U, suppose expo != 1
U_new = tmp./(ones(cluster_n, 1)*sum(tmp));
```

Distfcm_f is calculated as such

```
I=unique(data);
H1=[I,histc(data(:),I)];
H=H1(:,2);
tmp1=H.*I;

mf = U.^expo;        % MF matrix after exponential modification
center = mf*tmp1./(sum(mf,2)*ones(1,size(tmp1,1))*H); %new center
dist = distfcm_f(center, data);        % fill the distance matrix
obj_fcn = sum(sum((dist.^2).*mf));  % objective function
tmp = dist.^(-2/(expo-1));      % calculate new U, suppose expo != 1
U_new = tmp./(ones(cluster_n, 1)*sum(tmp));
```

3.3 Binarization

Binarization is the process by which a binary mask is applied over the image obtained on segmentation. Segmentation is the process of extraction of tumorous area from the input MRI image and is done using the proposed method (IFCM). On application of binary mask, each pixel takes the value of either 1 or 0. Value of the pixel is determined based on the input threshold. Hence, this process is also called thresholding. Let f be the input image and has pixel values in the gray scale range of k. Let threshold value be T, which also lies in the gray scale range k. Each pixel of f is compared with T (as shown in the Eq. 1). In case of the pixel value being less than T, the pixel is assigned a value of 0 and vice versa. This process acts as primary step in tumor size calculation.

$$g(n) = \begin{cases} 0 \;\text{ if } (n) \geq T \\ 1 \;\text{ if } (n) < T \end{cases} \tag{1}$$

where $f(n)$ is the transform coefficient,
T is the threshold value and
$g(n)$ is the transform coefficient after the binary decision for the same value of n.

3.4 Calculation of Tumor Size

In the approximate reasoning step, the tumor area is calculated using the binarization method. That is from the image having only two values either black or white (0 or 1). Here 256 × 256 JPEG image is a maximum image size. The binary image can be represented as a summation of total number of white and black pixels.

The tumor size is the summation of total number of white pixels. The users of the app will be provided various information depending on the tumor size. The relevant information for an appropriate tumor size will already be available in the app's database.

4 Implementation and Result

On the application of the clustering algorithm IFCM, the tumor cells are clearly segmented in the IFCM image, which is shown in Fig. 5. The various clustering algorithms are applied on the images in Fig. 4 which were obtained from dataset of brain tumor images provided by the department of computer science, University of Cyprus [9] and evident results are taken. A dataset of tumor-MRI images of the brain was obtained and the algorithm was tested on sufficient number of images (sample shown in Fig. 5). Figure 6a representing the IFCM iteration count takes negligibly small number of iterations whereas Fig. 6b corresponding to FCM takes around 42 iterations with a computation time of 290 s. In the denoised image, the edges are smoothened and now the image is ready for the application of IFCM. One of the sample's FCM and IFCM computation time is compared below. Figure 7 shows the clustered and segmented sample image using FCM and IFCM.

Image is segmented and thresholded by the proposed IFCM. We calculated the frequency matrix using histc function. Instead of multiplying with each data point, the distance matrix calculated was multiplied with the frequency matrix. Hence the number of iterations reduced significantly and the time complexity was also greatly reduced. The objective function was defined and in each iteration, it was reduced till the condition to exit out of the loop was satisfied. If the condition fails in the worst case, the iterations may run 100 times but still, the time complexity remains small.

As interpreted from Fig. 8, the computation time comes around 6.4719 s if the summation is taken over the column total time.

From Fig. 9, it is clear that the time of computation comes around 71.562 s for traditional FCM segmentation and thresholding. A speedup of 5.3 times is achieved by using IFCM.

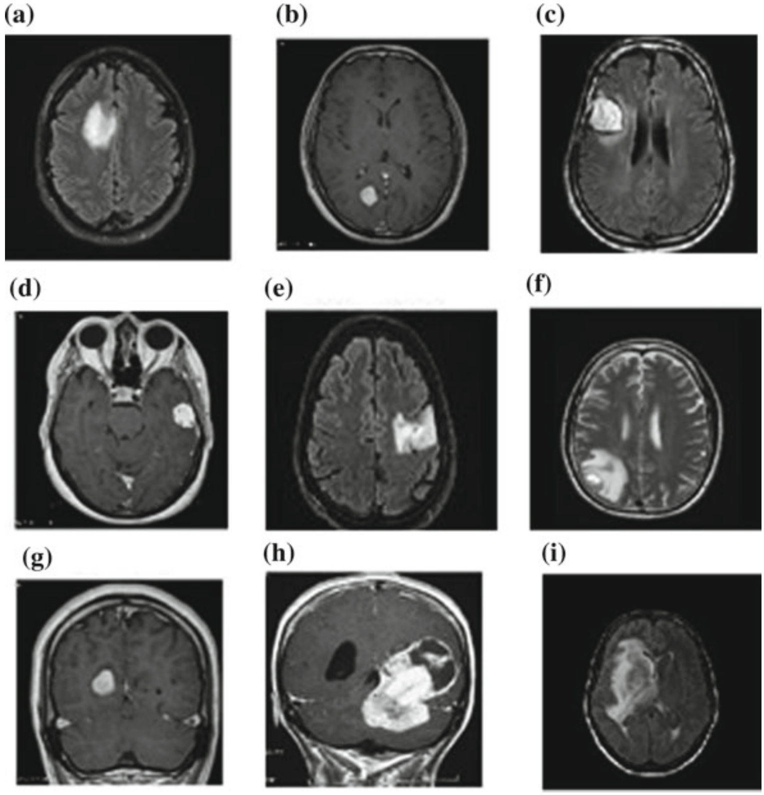


Fig. 4 Dataset of original images

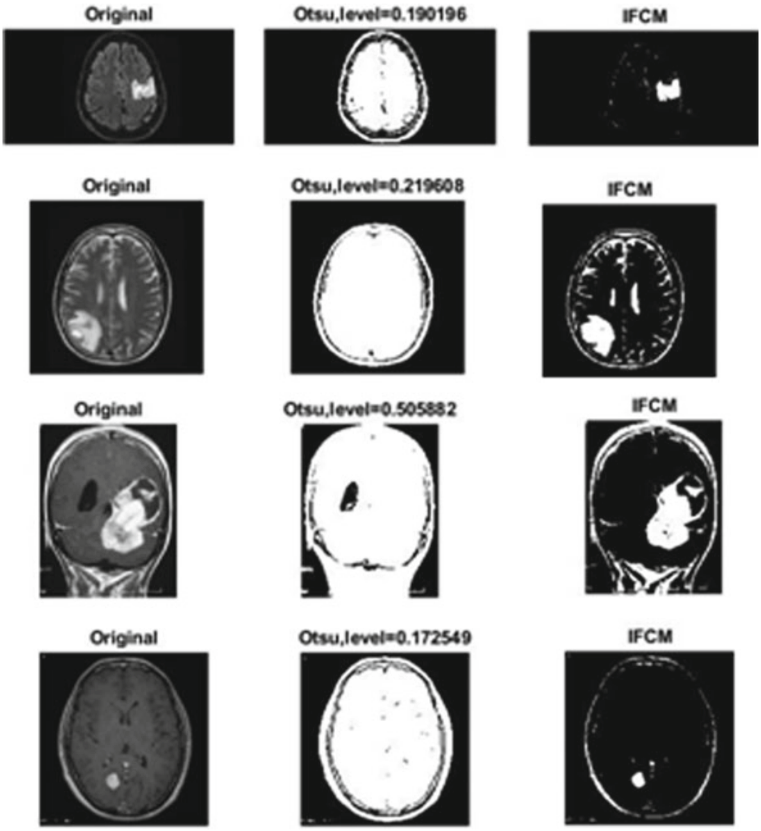


Fig. 5 IFCM samples

The computation time analysis for the best segmentation algorithms is discussed below. Their drawbacks are also mentioned with pictorial representations.

k-means algorithm is a hard clustering technique and hence a data point can belong to only one cluster, thus failing to give accurate results as shown in Fig. 10b.

Despite providing accurate results, FCM is a time-consuming process as the membership matrix of the size of the data matrix is calculated for each iteration. From the comparison table provided above, it is very clear that the computation time is nearly 10–20 times more than that of IFCM and k-means clustering techniques. Table 1 provides a comparative analysis of the dataset of 10 images in terms of output efficiency and computational time.

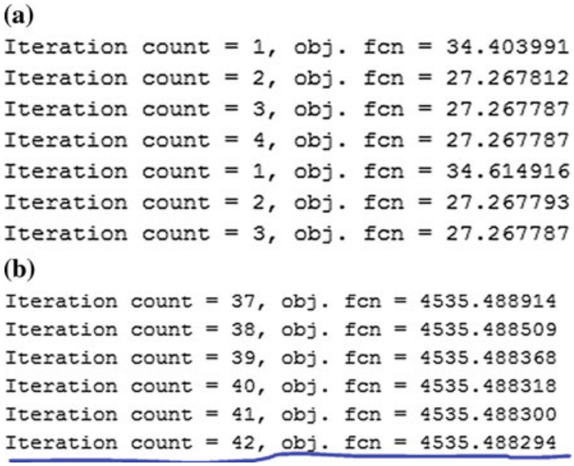

(a)
```
Iteration count = 1, obj. fcn = 34.403991
Iteration count = 2, obj. fcn = 27.267812
Iteration count = 3, obj. fcn = 27.267787
Iteration count = 4, obj. fcn = 27.267787
Iteration count = 1, obj. fcn = 34.614916
Iteration count = 2, obj. fcn = 27.267793
Iteration count = 3, obj. fcn = 27.267787
```
(b)
```
Iteration count = 37, obj. fcn = 4535.488914
Iteration count = 38, obj. fcn = 4535.488509
Iteration count = 39, obj. fcn = 4535.488368
Iteration count = 40, obj. fcn = 4535.488318
Iteration count = 41, obj. fcn = 4535.488300
Iteration count = 42, obj. fcn = 4535.488294
```

Fig. 6 **a** IFCM iteration, **b** FCM iteration

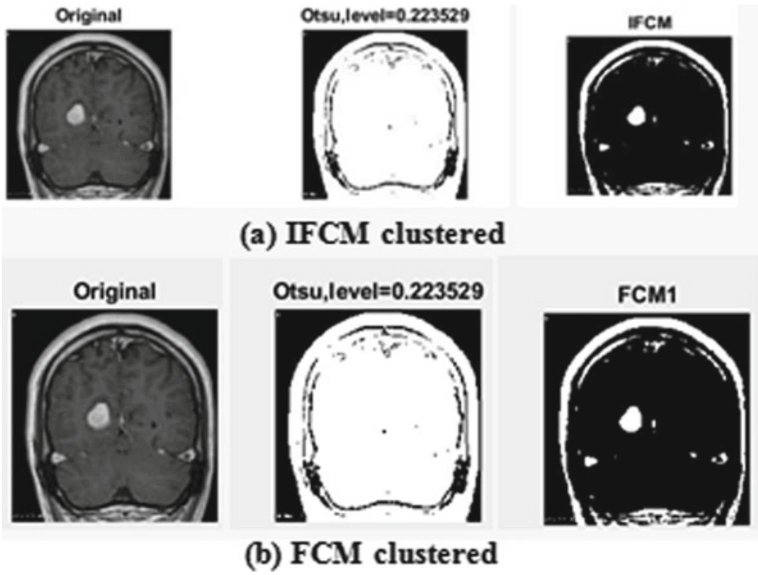


Fig. 7 Clustered images

Function Name	Calls	Total Time	Self Time*	Total Time Plot (dark band = self time)
m_f	1	1.438 s	0.008 s	
fcmthresh_f	2	1.080 s	0.005 s	
fcm_f	2	1.066 s	0.009 s	
stepfcm_f	7	0.945 s	0.290 s	
unique	16	0.763 s	0.004 s	
unique>uniqueR2012a	16	0.759 s	0.759 s	
distfcm_f	7	0.330 s	0.003 s	

Fig. 8 Time of computation for IFCM for sample Fig. 7a

Function Name	Calls	Total Time	Self Time*	Total Time Plot (dark band = self time)
M_ff	1	20.840 s	0.006 s	
fcmthresh	2	20.497 s	0.139 s	
fcm	2	20.354 s	0.018 s	
stepfcm	54	20.167 s	18.570 s	
distfcm	54	1.597 s	1.597 s	

Fig. 9 Time of computation for FCM for sample Fig. 7b

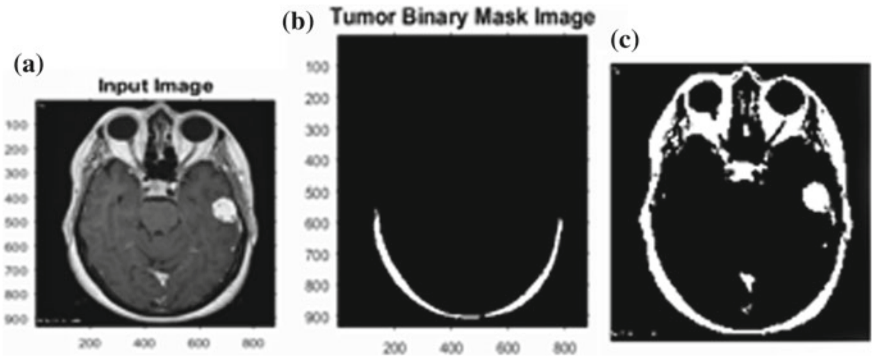


Fig. 10 **a** Original image, **b** clustered and thresholded image by k-means, **c** clustered and thresholded image by IFCM

5 Conclusion

In this paper, the performance metrics of our proposed algorithm IFCM is analyzed thoroughly in comparison with the preexisting algorithms. The IFCM proposed in our paper uses frequency in place of data points to provide evidently improved performance. A large number of images in dataset was computed in a faster and precise manner as possible. We believe that this conceptual proposal could serve for various image segmentation purposes, apart from serving the purpose of brain tumor recognition and analysis. Various high-resolution images were segmented in a very less time possible. Other major existing methods were also analyzed and drawbacks were also briefly discussed. Preexisting work's pros and cons were discussed and further chances of the hybrid models possible were also mentioned.

References

1. Hamerly G, Elkan C (2002) Alternatives to the k-means algorithm that find better clusterings. In: Proceedings on the 11th international conference on information and knowledge management (CIKM)
2. Lu Y, Chen W (2015) Automatic multimodal brain-tumor segmentation. In: IEEE fifth international conference on instrumentation and measurement, computer, communication and control
3. Pereira S, Pinto A, Alves V, Silva CA (2016) Brain tumor segmentation using convolutional neural networks in MRI images. IEEE Trans Med Imaging 35(5):1240–1251
4. Moon N, Bullitt E, van Leemput K, Gerig G (2002) Model-based brain and tumor segmentation. IEEE
5. Telrandhe SR, Pimpalkar A, Kendhe A (2016) Detection of brain tumor from MRI images by using segmentation and SVM. In: World conference on futuristic trends in research and innovation for social welfare (WCFTR-16)
6. Ghassabeh YA, Forghani N, Forouzanfar M, Teshnehlab M (2007) MRI fuzzy segmentation of brain tissue using IFCM algorithm with genetic algorithm optimization. IEEE
7. Forghani N, Forouzanfar M, Forouzanfar E (2007) MRI fuzzy segmentation of brain tissue using IFCM with particle swarm optimization. IEEE
8. Bezdek JC (1981). Pattern recognition with fuzzy objective function algorithms. ISBN 0-306-40671-3
9. www.medinfo.cs.ucy.ac.cy

An Off the Shelf CNN Features Based Approach for Vehicle Classification Using Acoustics

Anam Bansal, Naveen Aggarwal, Dinesh Vij and Akashdeep Sharma

Abstract Vehicle classification is a trending area of research in Intelligent Transport System. Vehicle Recognition can help traffic policy makers, public safety organizations, insurance companies, etc. It can assist in various applications like automatic toll collection, emissions/pollution estimation, traffic modelling, etc. Many methods, both infrastructure-based and infrastructureless have been proposed for vehicle classification but they have certain disadvantages. In this paper, we have explored the possibility to use off the shelf Convolutional Neural Network (CNN) features for commuter vehicle classification using acoustics. To extract features from acoustic recordings taken from the vehicle, a simple CNN is designed. These features are used to classify vehicles in five main categories car, bus, plane, train, and three-wheeler using Support Vector Machine (SVM). This approach is tested on dataset having 4789 recordings and gives good accuracy as compared to simple Mel Frequency Cepstral Coefficients (MFCC) feature based deep learning and machine learning approach.

1 Introduction

One of the challenging issues in Intelligent Transport System is to detect the mode of transport used by the commuters. It becomes more challenging on roads for developing countries like India where traffic is heterogeneous and chaotic. Vehicle Classification on roads can assist the traffic management policymakers. Knowing the type and the number of vehicles plying on road, various design decisions can be laid and layout of the respective city can be planned. Recognition of type of vehicle aid in the number of applications like automatic toll collection, surveillance, accident prevention, traffic congestion avoidance, etc.

A. Bansal (✉) · N. Aggarwal · D. Vij · A. Sharma
UIET, Panjab University, Chandigarh, India
e-mail: anambansal19@gmai.com

N. Aggarwal
e-mail: navagg@gmail.com

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_110

The researchers have proposed the number of sensing techniques to detect the type of vehicles plying on the roads. The proposed solutions can be broadly categorized into infrastructure-based and infrastructureless solutions. Infrastructure-based solutions include Magnetic Sensors, Inductive Loop Detectors, Video Cameras, Infrared Sensors, etc. These techniques have certain downsides. They are costly in terms of installation and maintenance. Further, the accuracy of video camera based approaches is affected by the weather conditions and occlusion. Similarly, magnetic sensor based approaches work well in homogeneous and lane driven traffic only. Even though, some researchers have proposed infrastructureless solutions, which mainly rely on smartphone sensors such as GPS, accelerometer, etc. [1]. But for vehicle classification, these approaches have very limited accuracy.

Various researchers have used acoustics for vehicle classification [2, 3]. Since each vehicle has a specific sound pattern, so they proposed to use the roadside installed microphones to capture the vehicular acoustics and extract their features for classification using SVM. Since it is difficult to install and maintain the roadside installed microphones, we propose to use acoustics captured from commuter's smartphone. Acoustics recorded are then sent to the server where CNN is used to extract features and SVM classifier is used to determine the type of the vehicle. Along with the commuter's acoustics, only current location is sent and no identity data such as phone number or ID are sent to preserve the privacy of commuters. Main contributions of this paper are:

1. Determining the efficacy of human-driven features and off the shelf CNN features using SVM classifier.
2. Tuning of various hyperparameters of SVM classifier.
3. Improving the vehicle classification accuracy using off the shelf CNN based approach as compared to human-driven features based approaches.

Remaining paper is structured as follows: In Sect. 2, survey related to vehicle classification is demonstrated. In Sect. 3, the system overview of the proposed approach is described. In Sect. 4, various experiments and their results are listed. In Sect. 5, conclusion and future work are described.

2 Related Work

Since vehicle classification is important in developing countries, various solutions have been proposed by researchers. Infrastructure-based solutions include Inductive Loop Detectors, Magnetic Sensors, Video Cameras, Infrared Sensors, Piezoelectric Sensors, Microwave Radars, etc.

Inductive Loop Detectors [4] and Wireless magnetic sensors [5] were used for classification of vehicles. But they are sensitive to traffic and temperature. Magnetic sensors need orderly and homogeneous traffic to perform accurately. Video cameras were employed for classifying vehicles [6]. But cameras, though can monitor several lanes but are affected by occlusion [7] and adverse weather conditions. Microwave radars were used for classifying vehicles into five classes [8]. But other electronic devices interfere with electromagnetic signals of radar. Overall infrastructure-based solutions are costly in terms of maintenance and installation. Then there are off road solutions for classifying vehicles such as GPS, GSM, and accelerometers [1, 9]. They are installed either in smartphones [10] or vehicles.

Sounds from the vehicles can be used for vehicle classification. Microphones capture the sounds emitted by the vehicles which can be used for determining various traffic parameters like presence, passage, and class of vehicles. Determining these parameters aid in applications like traffic congestion detection [11], accident detection [12], vehicle classification [13]. Traffic congestion state was detected by capturing sounds through microphones installed along the roads [14, 15]. Microphones in smartphones also give considerable accuracy for traffic state congestion detection [16, 17]. Microphones are not affected by lighting conditions and visual occlusions[11], work 24/7 and have low power requirements [16].

Microphones acquired the acoustic signals and classified vehicles into four classes—horns, medium, light, and heavy vehicles using Artificial Neural Network [3]. Vehicles' sounds were acquired by roadside installed microphones and detection and classification of vehicles were performed. MFCCs features were used and classification accuracies for Artificial Neural Network (ANN) and K-Nearest Neighbour (KNN) were compared [18]. Accuracies of Gaussian Mixture Model (GMM), Hidden Markov Model (HMM), and Bayesian Subspace Methods were compared for vehicle classification. Vehicles were classified into nine classes and MFCCs were extracted. Bayesian subspace method gave 50% higher accuracy as compared to others [2]. Probability Neural Network was used as the classifier for classifying vehicles' types and vehicles' positions [13]. Considerable accuracy was obtained. kandpal et al. [19] proposed feature extraction methods that helped in classification of ground vehicles on road in three categories—car, bike, and truck using Multilayer Neural Network. Military vehicles were classified into four classes by acquiring sounds from them. Accuracies of Multilayer Perceptron (MLP) and Probability Neural Network (PNN) based on Gaussian Mixture were compared. PNN gave higher accuracy [20].

We have proposed a new approach of inferring vehicle category through acoustic recordings of vehicles. CNN is employed for feature extraction and SVM for classification of sound recordings. The proposed approach surpasses the human-driven features based approaches for vehicle classification.

3 System Overview

The acoustic data from vehicles can be used to predict the type of vehicle. The vehicles on the roads generate various sounds such as friction between tires and pavement, gears, vibrations in the engine, rotational parts, wind effect, fans, etc. [2]. These sounds act as the signature for each vehicle. We have used off the shelf CNN features based approach. In this, CNN is used for feature extraction and SVM for vehicle classification. This approach is compared with the two approaches. In the

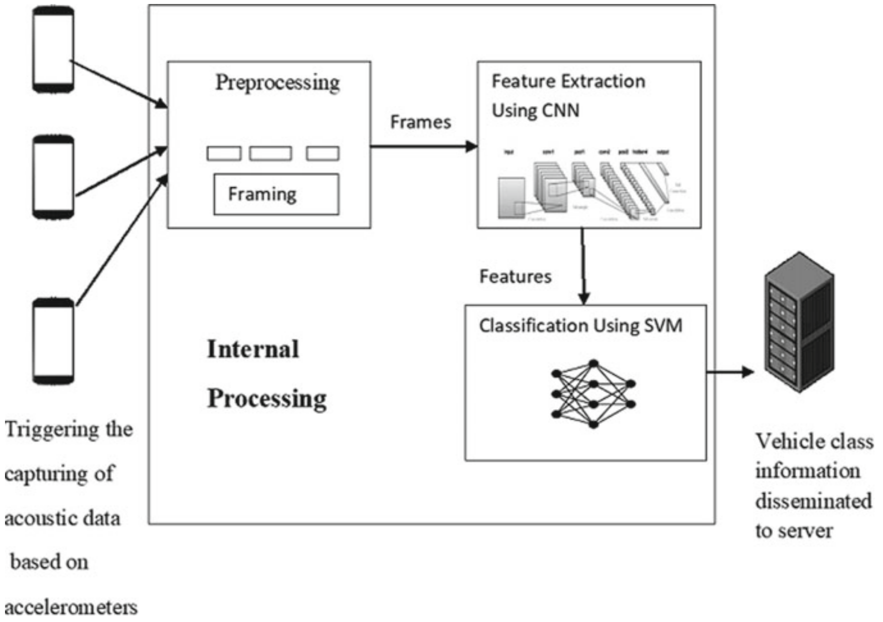


Fig. 1 System overview

first approach, MFCCs, extracted from audio signals are given as input to CNN and in the second approach, MFCCs are given as input to SVM. CNN and SVM classify the audio signals into one of the vehicles' categories.

The system overview of the proposed approach is shown in Fig. 1.

3.1 Preprocessing

In preprocessing phase, the audio signal undergoes framing. Framing is required so as to assume continuously varying signal quasi-stationary. Each audio signal is framed with the frame length of 8192. These frames undergo Discrete Cosine Transformation (DCT) and these coefficients are inputted to the CNN for feature extraction.

3.2 Feature Extraction and Classification

Features extracted from audio signals are the characteristics of vehicles that describe vehicle types. For off the shelf CNN features based approach, CNN is used for feature extraction.

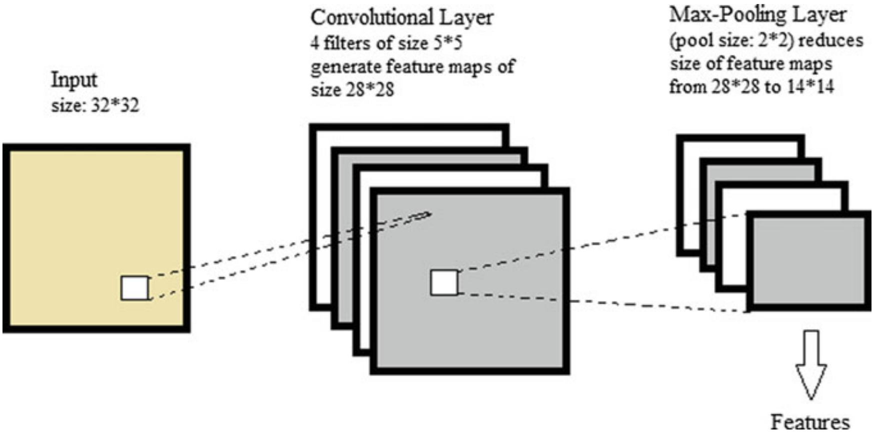


Fig. 2 Feature extraction using CNN

CNN has mainly four layers—Convolutional layer, Relu layer, Pooling layer, and Fully Connected layer. The last layer predicts the class of the vehicles. Before it, the layers extract features which act as the signature for respective class of vehicles. The convolutional layer has a number of filters. Each filter convolves across the audio signal. Each filter is trained to identify specific feature present in the audio signal. The result of filter's convolution is feature map (also called activation map). Set of activation maps become input to the next layer. Each successive layer identifies higher level features. We have used three layers of CNN—convolutional layer, Relu, and max-pooling layer. Features from max-pooling layer are given as input to SVM for classification. Figure 2 illustrates the feature extraction using CNN.

4 Experimental Setup and Results

The size of frame effects the accuracy of classification [16]. The frame size of 8192 is considered the best for traffic management applications. We have compared the accuracy of CNN, SVM, and the proposed off the shelf CNN features based approach. The number of experiments is conducted on the dataset consisting of 4789 recordings each of about 30 s [21]. Audio signals are sampled at 16 kHz sampling frequency. Each audio is classified into one of the five classes of vehicles—car, bus, train, plane, and three-wheeler.

The number of recordings for each class and the class labels are shown in Table 1. Each audio file undergoes framing with the frame length of 8192 and hop length of 2048. We experimented with training set size of 0.8, validation set size of 0.1, test set size of 0.1. The hyperparameters C and gamma are tuned to get the validation accuracy same as training set accuracy. C = 100 and gamma = 0.01 are the finally tuned parameters. Comparative analysis of three approaches is demonstrated in following subsections.

Table 1 Details of dataset

Vehicle category	No. of recordings	Class label
Car	850	0
Bus	1104	1
Plane	542	2
Train	1223	3
Three-wheeler	1070	4

4.1 MFCCs and SVM

Each audio signal undergoes framing and for each audio frame, 13 MFCCs features are extracted. These features are used to train the machine learning classifier, i.e., Support Vector Machine. We experimented with different kernel functions—Radial Basis function (RBF), Linear, and Sigmoid. RBF, Linear, and Sigmoid functions give accuracy of 72.65, 62, and 54.07%, respectively.

4.2 MFCCs and CNN

Each audio signal undergoes framing and for each audio frame, 13 MFCCs features are extracted and are inputted to Convolutional Neural Network. CNN is trained using categorical cross-entropy as the loss function. The size of mini-batch in each epoch is 32. The weights are updated using Adam after the mini-batch is processed. The first layer contains 256 neurons and input size is 13. Then there is another layer of 256 neurons and finally a fully connected layer. As an activation function, Relu is applied in between. There are 5 softmax nodes in the last layer corresponding to the number of classes. The CNN gives an accuracy of 79.97% on the test set.

4.3 Proposed Approach (CNN + SVM)

In this approach, raw audio signals undergo framing. Each frame of audio signal undergoes Discrete Cosine Transformation (DCT). The frame length of 8192 gives 8192 DCT coefficients. These 8192 coefficients are reshaped to the size of 48 * 48, interval wise. Though the other sizes of 36 * 36, 40 * 40, 44 * 44, and 52 * 52 were also tried but 48 * 48 yielded the best results. These are given as input to the convolutional layer of CNN. Input is convolved with filters in the convolutional layer. The dimensions of the output of convolutional layer are reduced by one max-pooling layer. The output from the max-pooling layer is flattened and given as input to SVM for vehicle classification. Experiments are performed with the different number of

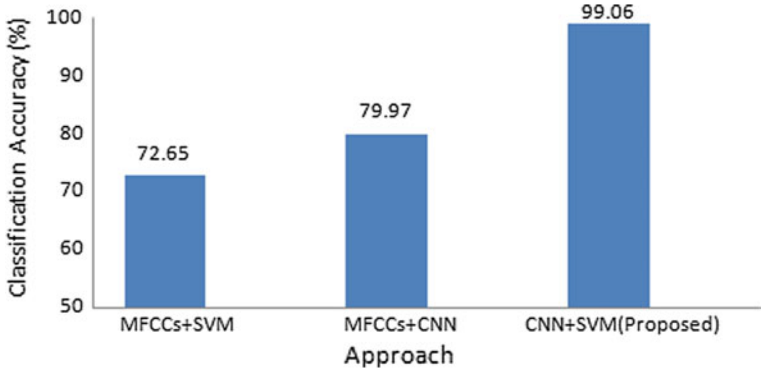


Fig. 3 Comparative analysis of three approaches

filters (4, 8, 16, 32), stride (1, 2, 4), and dimensions of filters (5 * 5, 7 * 7, 9 * 9), different pooling sizes(5 * 5, 7 * 7, 9 * 9). Figure 2 demonstrates the set of parameters for one of the experiments performed using this approach. Different kernel functions of SVM—Radial Basis Function (RBF), Linear function and Sigmoid function are used for experiments.

The best results are obtained with eight filters with filter dimension of 9 * 9, the stride of 2 and the pooling size of 9 * 9. RBF, Linear, and Sigmoid kernel functions give the accuracy of 94.89, 99.06, and 99.06% respectively.

Figure 3 illustrates the comparative analysis of three approaches.

5 Conclusion

In this paper, a new approach for vehicle classification using acoustics has been proposed. We have presented off the shelf CNN features based approach for classifying vehicles into five categories—car, bus, train, plane, and three-wheeler. The accuracy of the proposed approach is compared with SVM and CNN. It gave the accuracy of 99.06% which is quite higher than the approach using MFCCs and SVM (72.65%) and the approach using MFCCs and CNN (79.97%). We anticipate that the new approach can perform well in other acoustic domains as well by using transfer learning methodology and tuning various hyperparameters.

Acknowledgements This work is undertaken as a part of the project 'CARTS—Communication Assisted Road Transportation System' funded by ITRA, Media Lab Asia; and Design Innovation Center, Panjab University, Chandigarh, India.

References

1. Sun Z, Ban XJ (2013) Vehicle classification using GPS data. Transp Res Part C: Emerg Technol 37:102–117
2. Munich ME (2004) Bayesian subspace methods for acoustic signature recognition of vehicles. In: 2004 12th European signal processing conference, pp 2107-2110
3. George J, Cyril A, Koshy BI, Mary L (2013) Exploring sound signature for vehicle detection and classification using ANN. Int J Soft Comput 4:29
4. Gajda J, Sroka R, Stencel M, Wajda A, Zeglen T (2001) A vehicle classification based on inductive loop detectors. In: Proceedings of the 18th IEEE instrumentation and measurement technology conference, IMTC 2001, vol 1, pp 460–464
5. Kaewkamnerd S, Pongthornseri R, Chinrungrueng J, Silawan T (2009) Automatic vehicle classification using wireless magnetic sensor. IEEE international workshop on intelligent data acquisition and advanced computing systems: technology and applications, IDAACS 2009:420–424
6. Lai JC, Huang SS, Tseng CC (2010) Image-based vehicle tracking and classification on the highway. In: 2010 international conference on green circuits and systems (ICGCS), pp 666–670
7. Zhang W, Wu QJ, Yang X, Fang X (2008) Multilevel framework to detect and handle vehicle occlusion. IEEE Trans Intell Transp Syst 9:161–174
8. Urazghildiiev I et al (2002) A vehicle classification system based on microwave radar measurement of height profiles
9. Bhoraskar R, Vankadhara N, Raman B, Kulkarni P (2012) Wolverine: traffic and road condition estimation using smartphone sensors. In: 2012 fourth international conference on communication systems and networks (COMSNETS), pp 1–6
10. Singh P, Juneja N, Kapoor S (2013) Using mobile phone sensors to detect driving behavior. In: Proceedings of the 3rd ACM symposium on computing for development, p 53
11. Tyagi V, Kalyanaraman S, Krishnapuram R (2012) Vehicular traffic density state estimation based on cumulative road acoustics. IEEE Trans Intell Transp Syst 13:1156–1166
12. Ali HM, Alwan ZS (2017) Car accident detection and notification system using smartphone. LAP LAMBERT Academic Publishing, Saarbrucken
13. Paulraj M, Adom AH, Sundararaj S, Rahim NBA (2013) Moving vehicle recognition and classification based on time domain approach. Procedia Eng 53:405–410
14. Sen R, Raman B, Sharma P (2010) Horn-ok-please. In: Proceedings of the 8th international conference on mobile systems, applications, and services, pp 137–150
15. Sen R, Siriah P, Raman B (2011) Roadsoundsense: acoustic sensing based road congestion monitoring in developing regions. In: 2011 8th annual IEEE communications society conference on sensor, mesh and ad hoc communications and networks (SECON), pp 125–133
16. Kaur A, Sood N, Aggarwal N, Vij D, Sachdeva B (2017) Traffic state detection using smartphone based acoustic sensing. J Intell Fuzzy Syst 32:3159–3166
17. Vij D, Aggarwal N (2018) Smartphone based traffic state detection using acoustic analysis and crowdsourcing. Appl Acoust 138:80–91
18. George J, Mary L, Riyas K (2013) Vehicle detection and classification from acoustic signal using ANN and KNN. In: 2013 international conference on control communication and computing (ICCC), pp 436–439
19. Kandpal M, Kakar VK, Verma G (2013) Classification of ground vehicles using acoustic signal processing and neural network classifier. In: 2013 international conference on signal processing and communication (ICSC), pp 512–518
20. Maciejewski H, Mazurkiewicz J, Skowron K, Walkowiak T (1997) Neural networks for vehicle recognition. In: Proceeding of the 6th international conference on microelectronics for neural networks, evolutionary and fuzzy systems, p 5
21. Aggarwal N, Vij D, Soni S (2017) Acoustic vehicular data. Retrieved from http://pudataset.puchd.ac.in:8080/jspui/handle/123456789/14. Accessed on 10 Apr 2018

Conjunctival Vasculature Liveness Detection Based on DCT Features

S. N. Dharwadkar, Y. H. Dandawate and A. S. Abhyankar

Abstract Iris liveness detection algorithms are developed to recognize iris images were acquired from a live person who is actually present at the time of data capture. However the quality of the acquired images will decide success rate. Systems can be spoofed by using fake Photographs, video recordings, printed contact lenses, etc. Conjunctival Vasculature can be used as a biometric trait to identify liveness, paper gives focus on generation of a novel method to extract significant portion of off-angle eye called as sclera. DCT Transform based statistical features are used to find liveliness. System is tested using Extreme learning machines.

1 Introduction

In the past decade, there are found many applications based on biometric-based personal authentication technologies. Biometrics is the study of personal identification using highly unique behavioral characteristics of human beings. Biometric systems have proved their accuracy and convenience of use of requirement for better and robust security in our interconnected world. Liveness detection is measure concern for all security systems worldwide. It is used to find data is taken from live person or not. This is especially crucial for remote authentication. Number of biometric traits has been used to check liveliness [1]. Among all biometrics Conjuctival vasculature is most recent one which can be used along with IRIS.

Conjunctival vasculature recognition is most recent ocular biometric modality. It can be captured in the visible spectrum which is possible in all camera devices.

S. N. Dharwadkar (✉) · Y. H. Dandawate (✉)
E&TC Department, VIIT, Pune 48, India
e-mail: shridharwadkar@gmail.com

Y. H. Dandawate
e-mail: yhdandawate@gmail.com

A. S. Abhyankar (✉)
E&TC Department, SPPU, Pune, India
e-mail: aditya1210@gmail.com

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_111

Simona Crihalmeanu and Arun Ross explained conjunctival vasculature can be used as a soft biometric [2]. Due to rich layer of blood vessel pattern this trait cannot be spoofed easily hence can be used as an assisted technique for robust iris recognition and checking liveliness. They provide detailed explanation for an extraction of conjunctival vasculature pattern from multispectral eye images. After minutiae extraction HU invariant moments were calculated. Identification was done on the basis of correlation. Accuracy was obtained 100% but experimentation was being carried on Limited database of 49 objects with manual process used for segmentation. Seyed Mohsen Zabihi, Hamid Reza Pourreza et al. explained use a classifier-based method for segmentation and extraction of blood vessels from conjunctival images [3]. In their proposed work, Contrast Limited Adaptive Histogram Equalization (CLAHE) method was used to distinguish vessels from background as a preprocessing task. It also explains further, extraction of feature vector by using Local Binary Patterns (LBPs) and finally Artificial Neural Fuzzy Inference System (ANFIS) classifier is used for segmentation of pixel vessels with background. Morphological operations were used as a post processing step applied to improve the results of segmentation process along with removal of noise. However algorithm was tested on 40 conjunctival images. Number of researchers makes use of Iris for liveness detection. Kang R young Park et al. introduced the novel method for fake iris detection [4]. They tracked changes in dimensions of pupil along with local iris features in presence of visible light. Also to detect the change of local iris features, multiple wavelet filters having Gabor and Daubechies bases were used. To estimate accuracy of fake iris, SVM (Support Vector Machine) were used. Virginia Ruiz-Albacete and Pedro Tome-Gonzalez explain in their excellent review discussed emergence of various ambiguities due to direct attacks which of iris-based verification. Fake database was created by printing iris images by using commercial printer and then, presenting images in front of iris sensor [5]. He also highlights different attacks like effect of light reflections or eye movement, pupil response to a sudden lighting event on recognition. Huang, Changpeng Ti, Qi-zhen Hou, et al. have done An experimental Study of Pupil Constriction for checking liveliness [6]. They made use of pupil constriction, to differentiate real and live irises. Vikas Gottemukkula, Sashi Kanth Saripalle et al., used fusion of iris and conjunctiva for recognition [7]. They design a weighted fusion scheme to combine IRIS and conjunctiva. Experiments were conducted on 50 persons indicate that such a fusion scheme gives reduction in equal error rate by of 4.5% as compared with iris recognition.

IRIS-based system can be spoofed by using direct attacks so to add robustness, conjunctival vasculature based modality can be used along with iris. As this modality works in visible spectrum along with robustness it is gaining more and more popularity. Section 2 explains methodology of Fake database generation. It also highlights novel algorithm developed for automatic Sclera extraction from database image. Section 3 provides analytical discussion of a liveliness detection system in terms of statistical features based on DCT transform. Section 4 provides concluding remarks and future scope.

2 Proposed Methodology

In a Conjunctival vasculature based liveness detection an eye image entered in the system can be differentiated as real or fake, so the extraction of image features from region of interest is most important task. Algorithm can be summarized as follows.

2.1 Image Acquisition for Database Generation and Experimentation

Liveness detection based on conjunctiva has found limited exposure in literature. So there is need for generation of an off-angle fake iris database [8]. Database of 719 off-angle IRS images is generated out of which 600 are real images. Out of that 60% are male volunteers and remaining are female volunteers with age group from 20 to 35 years. Figure 1 provides samples of Off-Angle Iris images. Off-Angle eye images

Fig. 1 Original eye image

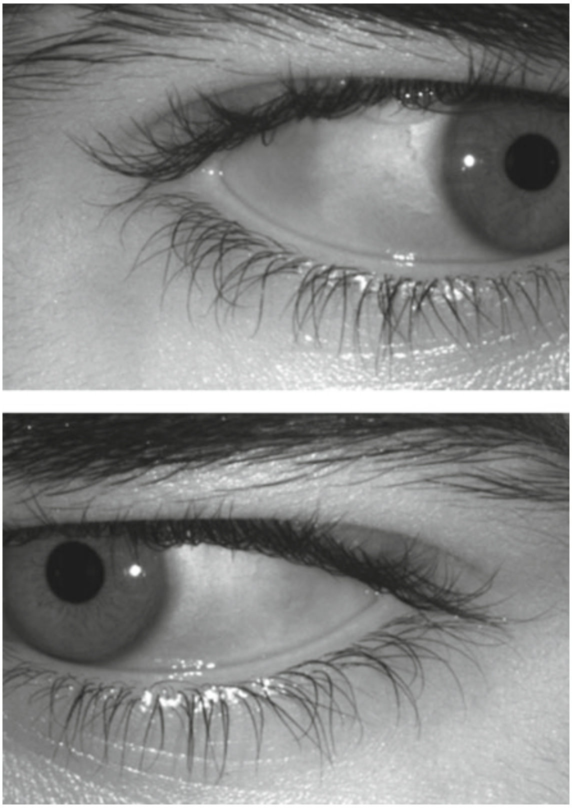


Fig. 2 Fake eye image

are captured by using CMITech scanner is a binoculars-type iris biometrics imaging device with Iris image pixel resolution of 640 × 480 pixels. As conjunctival features can be capture in visible spectrum fake database is created by taking photograph by 11 megapixel Sony digital camera. Figure 2 shows sample of fake database images.

A ROI Extraction

After image acquisition significant portion of an eye called as conjunctival vasculature needs to be separate out from as eye image. Proposed preprocessing algorithm follows following steps.

(1) Capture an eye image.
(2) Binarize Input Image by using thresholding.
(3) Apply Adaptive Histogram Equalization.
(4) Crop eye image by drawing Concentric circles with radii equal to that of size of an Iris. Integro-differential operator proposed in [9] will be used for iris localization.
(5) Find different areas of connected components in the binary Image.
(6) Find Maximum area component which extracts white portion of an eye.
(7) Prepare Binary mask.
(8) Apply it on original image to extract sclera pattern.
(9) Fit maximum square from extracted ROI.

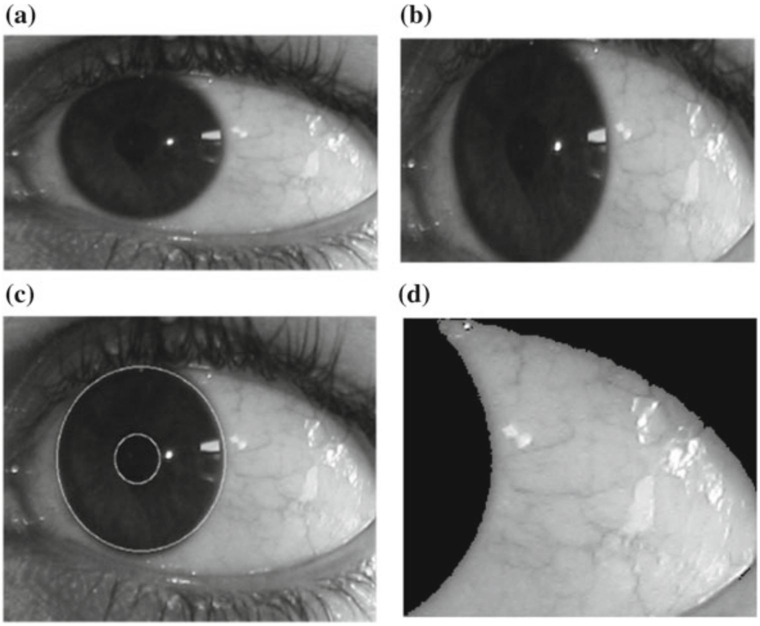


Fig. 3 **a** Original eye image, **b** iris localization figure, **c** cropping based on iris localization, **d** ROI extraction

B Feature Extraction

After extraction of Maximum Square from extracted ROI, DCT transform was applied to highlight low-frequency coefficients compare with high frequency components. DCT transform can be calculated as shown in Eqs. 1–3 respectively.

$$C_k = \frac{2}{N} w(k) \sum_{n=0}^{N-1} x_n \cos\left(\frac{2n+1}{2N}\pi k\right), 0 \leq k \leq N-1 \tag{1}$$

and

$$x_n = \sum_{k=0}^{N-1} w(k) C_k \cos\left(\frac{2n+1}{2N}\pi k\right), 0 \leq n \leq N-11 \tag{2}$$

where

$$w(k) = \sqrt{2}, k = 0 \text{ and } w(k) = 1, 1 \leq k \leq N-1 \tag{3}$$

The use of DCT can produce better results with less computational cost Statistical features like mean, Standard deviation, Variance and Mean were calculated from transform coefficients. First order features can be calculated as follows:

$$\text{Mean} = \frac{1}{mn} \sum_{(r,c)\in W} g(r, c), \tag{4}$$

where "g" is the image, "r" and "c" are the row and column coordinates respectively, within image size of "$m \times n$"

$$\text{Variance} = \frac{\sum f.\left(X_1 - \bar{X}\right)^2}{N} \tag{5}$$

$$\text{Standard Deviation} = \sqrt{\frac{\sum f.\left(X_1 - \bar{X}\right)^2}{N}}, \tag{6}$$

where $\bar{X}$ is mean and N is size of an image

$$\text{Median} = \text{argmin}_n,\ \alpha(H(n) - |n - \alpha| \tag{7}$$

Figure 3 highlights different steps used for extraction of conjunctiva.

C **Extreme learning Machine Approach**

Extreme learning machine is used to classify liveliness of an image. ELM is Single-layer feed forward neural network which provides much faster speed than traditional feed forward network learning algorithms. Advantage lies in assigning random input weights and hidden layer biases of SLFNs [10]. After the input weights and the hidden layer biases are chosen randomly, inverse operation of the hidden layer output vector can then be used to determine the output weights (linking the hidden layer to the output layer) of SLFNs.

3 Experimental Results

Experimentation was carried on 719 images with 600 live and 119 fake images Cross-validation ($k = 2$, 5- and 10fold) is performed to calculate accuracy of the system. Mean, Standard Deviation, Variance and Median were used as a member of feature vectors for training and testing of the system. Sigmoid activation function with 10 numbers of neurons is used for training for optimum performance. Table 1 results proved that system performs better for 10fold cross-validation. Experimentation carried out on Matlab R2015a software with Intel i5 processor with 1.7 GHz speed. Table 2 shows Percentage accuracy of discrete cosine transform with all features for ($K = 2/5/10$) fold cross-validation.

Figure 4 shows calculation of equal error rate (EER) from Receiver Operating Characteristic (ROC) plot. EER is measured at intersection of $-45°$ line with the characteristics. EER with all feature vectors found to be 0.1. Area under Curve (AUC) is 0.987 provides good quality of Classifier. Table 2 shows accuracy of system for different combinations of feature vectors with 10fold cross-validation.

Table 1 Percentage accuracy of DCT transform for cross validation

Transform/method	Features	Percentage accuracy		
DCT Activation function:Sigmoid No. of neurons = 10	Mean + std. deviation + variance + median	$K = 2$	$K = 5$	$K = 10$
		93.88	94.02	94.02

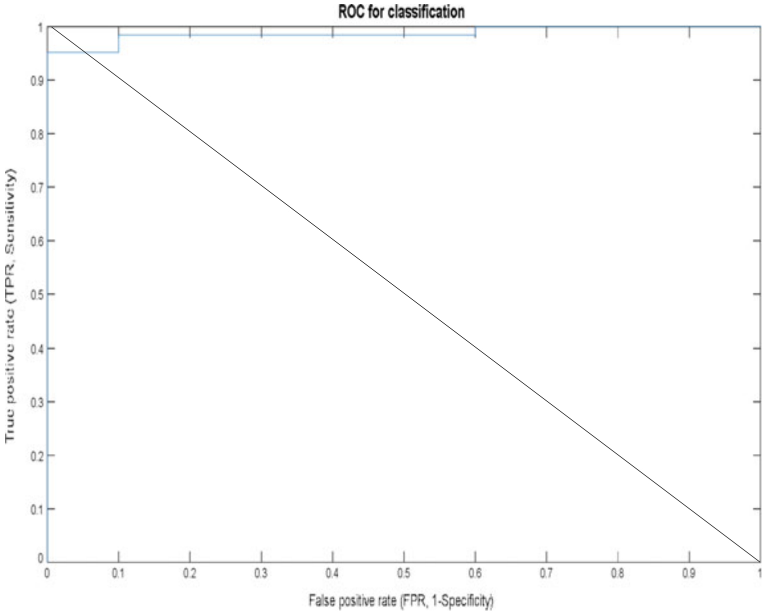


Fig. 4 ROC plot

Table 2 Percentage accuracy of DCT transform for various combination of features

Features	Percentage accuracy
Mean	90.13
Std. deviation	87.21
Variance	90.54
Median	83.59
Mean + variance	93.88
Mean + std. deviation	93.88
Variance + median	86.65
Mean + std. deviation + variance	93.88
Std. deviation + variance + median	86.23
Mean + variance + median	94.02
All features	94.02

4 Conclusion

Paper provides preliminary work based on use of Conjuctival vasculature for liveness detection. It presents novel technique for Sclera extraction from an off-angle eye image. Robustness of system demands reduction of equal error rate. Further investigation is required which transform provides maximum accuracy. Performance of the system needs investigation depending on environmental conditions at the time of database creation.

References

1. Schuckers S (2002) Spoofing and anti-spoofing measures. Inf Secur Tech Rep 7:56–62
2. Crihalmeanu S, Ross A (2011) On the use of multispectral conjunctival vasculature as a soft biometric. In: Proceedings of IEEE workshop on applications of computer vision (WACV), Kona, USA, Jan 2011, pp 204–211
3. Zabihi SM, Pourreza HR, Banaee T Vessel extraction of conjunctival images using LBPs and ANFIS. Int Sch Res Netw ISRN Mach Vis 2012(Article ID 424671)
4. Park KR, Whang MC, Lim1 JS, Cho Y (2006) Fake iris detection based on multiple wavelet filters and hierarchical SVM. Springer, ICISC 2006, LNCS 4296, pp 246–256
5. Ruiz-Albacete V, Tome-Gonzalezm P (2008) Direct attacks using fake images in iris verification. Springer, BIOID 2008, LNCS 5372, pp 181–190
6. Huang X, Ti C, Hou Q-z (2013) An experimental study of pupil constriction for liveness detection. In: IEEE workshop on applications of computer vision (WACV), pp 252–258
7. Gottemukkula V, Saripalle SK (2012) Fusing iris and conjunctival vasculature: ocular biometrics in the visible spectrum. In: IEEE conference on technologies for technologies, pp 150–155
8. Schuckers SAC, Schmid NA, Abhyankar A (2007) On techniques for angle compensation in nonideal iris recognition. IEEE Trans Syst Man Cybernetics Part B 37(5):1176–1190
9. Daugman J (1993) High confidence visual recognition of persons by a test of statistical independence. IEEE Trans Pattern Anal Mach Intell 15(11):1148–1161
10. Huang G-B, Zhu Q-Y, Siew C-K (2006) Extreme learning machine: theory and application. Neurocomputing 70:489–501

Unusual Social Event Detection by Analyzing Call Data Records

V. P. Sumathi, K. Kousalya, V. Vanitha and N. Suganthi

Abstract The availability of call data records provides an opportunity to identify unusual social events occurring in the society in an effective manner. The visitors participating in the events are the most important stakeholders for event organizers to improve their success rate. Visitors global positioning system (GPS) enabled device provides spatial data that are used to identify the visitors' presence during event time. Mobile call data records (CDR) represent a real spatial data source to detect occupied visitors, but provide less accuracy in terms of time and space resolution. Using spatial data, it is possible to detect unusual events. In this paper, the method for detecting unusual social events in various locations is proposed. The CDRs are used to detect new visitors who participated in the event and total number of visitors participated is computed for free-to-view social event. The information extracted from preprocessed CDRs are utilized to identify new visitors and to compute total visitors present in an event place effectively. The tower-wise visitor's details and CDR details provide information about the visitors' movement as well as CDRs distribution pattern during the event time. The experiment is conducted on real CDRs provided by a telecommunication service provider (TSP) servicing in a larger city. Results show that the proposed method provides accurate identification visitors involved in unusual social events compared to the state-of-the art methods.

V. P. Sumathi (✉) · V. Vanitha · N. Suganthi
Department of Computer Science and Engineering, Kumaraguru College of Technology, Coimbatore, Tamil Nadu, India
e-mail: sumathi.vp.cse@kct.ac.in

V. Vanitha
e-mail: vanitha.v.cse@kct.ac.in

N. Suganthi
e-mail: suganthi.n.it@kct.ac.in

K. Kousalya
Department of Computer Science and Engineering, Kongu Engineering College, Perundurai, Tamil Nadu, India
e-mail: kouse@kongu.ac.in

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_112

1 Introduction

In the past 5 years, mobile users and their societal-related activities have been increasing enormously. In developing countries, almost 70% of the people have smart mobile phones and have been participating in chatting and messaging all the time. A CDR is a record of a call setup and completion, and its format varies among telecom providers or programs, some of which are allowed to be configured by the user. It is also known as call detail record. Apart from mobile users' call and SMS data, the TSP creates data related to mobile network and mobile user profile. Naturally, the CDRs are in large volume and have privacy and security. Analysis of these data helps in deriving meaningful information like people's commuting behavior within a city [15, 17], inferring attractiveness towards public events and significance of social trends. It also helps in city planning [9, 14], marketing, customer profiling, food consumption patterns, disease spreading patterns, people's response to armed conflict, natural disasters, social event occurrence and its impact.

There are two kinds of events: (i) pay-to-view where the details of the event are already well known by organizer and these events are pre-planned. The number of participants, event duration and event location are fixed during the planning time, and (ii) free-to-view event where the participant's details may not be available until the day of the event, but the location and event duration are fixed. The free-to-view social event has to be analyzed to understand various factors like participant count for efficient crowd management, event planning and resource prioritization. In addition, understanding of user mobility patterns and events of interest [7, 8] are also of significance. This offers support for traffic and road management.

One of the important benefits of the social event analysis is identifying more involved visitors, who are very important for measuring the event success of an event organizer. The success can be measured using direct results like visitor count, involvement of the visitor and indirect factors like feedback from visitor. To understand the details of a visitor, it is necessary at first to find the retention of the existing visitor, the total count of the visitor, the visitors' mobility during the event time and the visitors' involvement in the event.

2 Review of Literature

CDRs contain both spatial and temporal data, facilitating the study of the mobility patterns and social behavior of individuals as well as the crowd. This rich source of data has also been taken into consideration to understand the success of an event by estimating the visitor count. Conventional methods of estimating the participant count like sampling techniques to estimate the number of participants or count the number of feedback collected may be biased. Z-score based event attendance estimation approach has been proposed [4, 10] with median error of less than 15%. It first identifies the towers associated with the event location and uses the CDRs generated by the network cells identified during the event to estimate the number of attendees. During the event, the total calls and SMS made by the mobile users are found to be directly proportional to the crowd size in that area [1, 3]. CDRs have been utilized to measure the effectiveness of billboard advertising. A system that estimates the effectiveness of an advertising based on the number of mobile phones near the billboard and infers people's preferences by using freely available event information on the internet is proposed [12].

The occurrences of social events like strikes or riots have been studied using mobile data [5]. The Bayesian location inference framework is proposed to detect unusual social gathering and to indicate the users who would have attended the event, and when and where an event has happened. Probabilistic approach has been employed to take into account the erratic antenna jumps [16]. The user mobility patterns are analyzed and the origins of people attending an event have been shown to be strongly correlated to the type of event. In addition, predictive analytics has been used to understand the users' interest patterns [2, 11]. The digital footprints created through mobile phones are utilized to rank the attractiveness of the place [6]. The social event recommendation system using K-nearest neighbor (K-NN) algorithm has been designed using mobile phone data [13]. Solving crowd estimation problem can help event managers predict the success of an event, estimate the number of participants and identify participant cluster which is engaged with mobile phone during the event time. As the CDR generated depends on the number of visitors in the event location, the problem of bias can be avoided.

2.1 Motivation

The present paper focus on analyzing the large amount of CDRs generated by telecommunication equipment, which is used to make decision to identify the unusable social event, to analyze the behavior of the visitors involved in the event, to identify the tower-wise CDR cluster and visitor count engaged mostly with mobile phone.

3 Materials and Methods

Social events are analyzed using CDR attributes. The six attributes related to social behavior analysis are total calls, number of incoming calls, number of outgoing calls, calls made to individuals physically located in the city, number of unique contacts, fraction of contacts called in the previous month and so on. The current study makes an attempt to identify unusable free-to-view events held at various locations in a city. The challenge is that the event location considered is highly distributed and the duration of the event is more than 7 h. To find out the visitor count, the amount of time spent by an individual at the event is considered. To get an accurate number of visitors, the CDRs of people who have their home or work base in that particular event location are removed.

3.1 Data Set

The CDRs are collected from a mobile service provider of Bengaluru, a city in India. The input data set contains CDRs for one month for both call and SMS. The call type attributes contains four types of values like 0, 1, 2, and 3. The values 0, 1 represent incoming and outgoing calls while 2, 3 indicate incoming and outgoing SMS. The total call duration is in the form of seconds. IMSI is used to identify the presence of other state visitors in the event while IMEI is used to identify the visitors with more than one Subscriber Identity Module (SIM). The call initiation and termination are identified using starting tower identity and ending tower identity numbers. The call status indicates success and failure of the call or SMS. Network cell details are utilized to know about the tower details like location of the tower and unique tower identity. The input tower data contain more than 10,000 tower details. All these towers are erected within Bangalore city. The latitude and longitude are used to compute the radius of an event area. The distance is calculated by using the formula as in Eq. (1).

$$\begin{aligned} \text{Calculated distance (in Kilo Meter)} = {} & A\text{COS}(\text{COS}(\text{RADIANS}(90 - \text{Lat1})), \\ & \times \text{COS}(\text{RADIANS}(90 - \text{Lat2})) \\ & + \text{SIN}(\text{RADIANS}(90 - \text{Lat1})) \\ & \times \text{SIN}(\text{RADIANS}(90 - \text{Lat2})) \\ & \times \text{COS}(\text{RADIANS}(\text{Long1} - \text{Long2})) \times 6371 \end{aligned} \quad (1)$$

where Lat1 and Lat2 denotes latitude and Long1 and Long2 denote longitude. The system is implemented using Python language, Graphlab with Turi machine learning environment.

3.2 Preprocessing of CDRs

The mobile CDRs observed by different mobile service providers do not have uniform format and structure. The values observed are large in size. Almost 50 million CDRs are generated every day. There is a need to process the CDRs generated and to convert them into a suitable form for applying data analytics. The data preprocessing step is quite important in any data mining application to achieve higher accuracy. The pre-processing involves missing value handling and outlier detection.

3.2.1 Missing Value Handling Techniques for Mobile CDR

The collected CDRs may contain missing value records due to system malfunction, network failure, mobile phone failure, power failure, and so on. For estimating visitor count, the missing value CDRs are identified and removed. This missing value removing technique is termed as reduction of data set. This technique is not suitable in all aspects of CDR analysis. The important task is to identify the unique users present in event location. The missed value record removal does not provide much effect on visitor estimation. The mobile user, making a call or SMS, is enough to identify the unique visitor in that location.

3.2.2 Outlier Detection and Removal from CDR

CDR needs to be pre-processed to remove the outliers. Outliers are mobile users who make a huge number of calls or SMS than the normal mobile users. The outlier call details are identified by counting the total number of CDRs generated by each user per day. If the total number of CDR associated with one particular user is greater than the specified threshold value, then that user will be considered as outlier. The threshold is selected by analyzing the call distribution.

3.2.3 Analyzing New Visitors' Presence in a Specific Location

Visitors' presence during the event is analyzed by first identifying the presence of regular users in the event location other than event day, and then removing these users from the CDRs generated in that event location during the event. This step is repeated for several days before and after the event day. The mobile users, who make and receive at least one call or SMS every day, are considered as regular users. It is understood that these users are living, working or commuting in the event location other than the event day and event time. The remaining CDRs are associated with the presence of new visitors in the event area during event time and the same helps in estimating the visitor count.

3.2.4 Analyzing Visitors' Presence During an Event Time

During the event day, the presence of visitors for a fraction of time at the time of the event is directly proportional to the presence of the visitor in the event. In the same way, the fraction of time in which the visitor is there outside the event time is inversely proportional to the presence of the visitor in the event. The inter-CDR time is calculated by considering time difference between the first CDR generated and second CDR generated on the same day. To identify the amount of time spent by the user in the event location, Maximum Inter-CDR Time (MICT) is computed by considering first and last call or SMS made by the visitors. MICT is the time difference between the first CDR and the last CDR generated on the same day. It is computed for every visitor in the event location on the day of the event. Its value is used for the estimation of visitor count. The threshold value is fixed at 60% based on the events like sports, conferences, cultural, social events, and so on and the event location whether it is distributed or centralized. The CDRs with MICT value less than the threshold value are considered as the normal mobile users.

3.3 Tower-Wise Visitor Clustering Algorithm

In usual events, the actual participant is known by counting the number of tickets sold on event day, which is called ground truth. However unusable events, the visitor count is not always equal to ground truth because (i) some visitors may not make any call or SMS during the event day and (ii) all the visitors may not belong to the same service provider. The visitor clustering algorithm is given in Fig. 1.

3.4 Tower-Wise CDRs Clustering Algorithm

During unusual events, the CDR count gets either increased or decreased based on the nature of the event. The CDR count always depends on the number of visitors involved in the event day in that location. The tower-wise CDR clustering is used to understand the visitors' behavior belonging to the same service provider as given in Fig. 2.

Visitor clustering algorithm (CDRs and tower location details)

Input: CDRs filtered based on date, location, duration of event and tower identity
Output: Tower-wise participant cluster.

1. Extract both call and SMS CDRs associated with each mobile user.
2. Combine both call and SMS CDRs associated with a particular mobile user.
3. Find the time of first CDR generated during a particular day.
4. Find the time of last CDR generated during a particular day.
5. Compute the time difference between first and last CDR which is equal to MICT of visitor.
6. The visitors count is computed based on MICT value of individual user.
7. The visitors are clustered based on tower usage pattern.

Fig. 1 Visitor cluster algorithm

CDR clustering algorithm (CDRs and Tower location Details)

Input: CDRs filtered based on date, location, event duration and tower identity
Output: Tower-wise CDR cluster.

1. Extract both call and SMS CDRs associated with specific location.
2. Combine both call and SMS CDRs associated with all mobile users.
3. Identify and remove all outliers as well as regular users. Consider only the new user present in the event.
4.The visitor count is computed based on MICT value of individual user.
5. The identified visitors' CDRs are filtered and considered for clustering.
6. Cluster the CDR based on date, time and tower usage details.

Fig. 2 CDRs clustering algorithm

4 Result and Discussion

By nature, CDRs are in large volume (Bid data), completely raw (not suitable for data mining), pattern extraction is complicated, real-time data and useful features must be identified. Limitations of CDRs are call frequency that depends on wealthier people, zero tower issue while non-mobile users are excluded. The users considered are educated who lives in urban areas. The pit falls of CDRs are high density of towers in cities, lower density of towers in rural area and frequent users versus new users. The users are making more than 100 calls or SMS considered as outliers. After removing the outliers, the CDR distribution pattern throughout the day is shown in Fig. 3. The data of all the days, except March 10, 2013, are obtained and found to be of similar pattern of distribution. Minimum of 99.50% of mobile users are considered as normal user. Similarly the call distribution patterns of all Sundays in March 2013 are compared in Fig. 4. In both the graphs, March 10 shows dissimilar pattern compared to all the other days. Some unusual events that occurred on that day are clearly shown in the graphs.

Towers present within 1 km (30 towers) in and around the event location are considered. This place always has more people during the night time. The event time duration from 6.00 p.m. to 12.00 p.m. is considered for counting new visitors in that location. The visitors count was more on all Sundays 2013, except March 10th as shown in Fig. 5. It clearly shows that some unusual things have occurred on that day.

Table 1 shows the new visitor count on all Sundays. The days considered are 3, 10, 17, 24, and 31 of March 2013. The users present in all the days (intersection) are considered and their CDRs are removed while those present in any one of the days (union) other than the day considered (March 10) are removed. The new visitor count is 454 (intersection) and 219 (union) on March 10. Some unusable events could occur on March 10.

After identifying the visitor count, the visitors are clustered based on their tower usage pattern during the event day and time. Figure 6 clearly shows the tower-wise

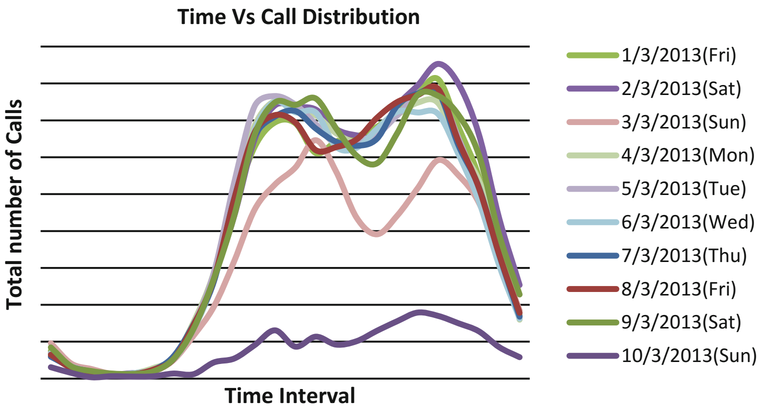


Fig. 3 Call distribution for 10 days

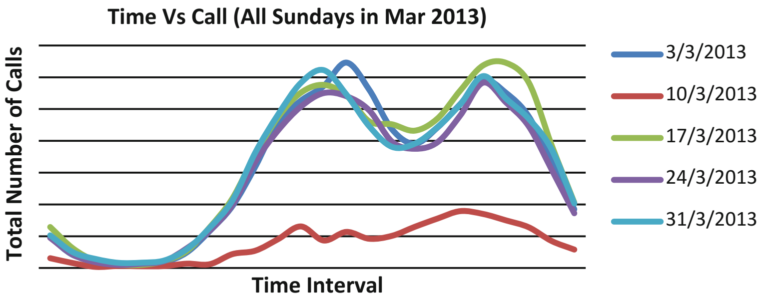
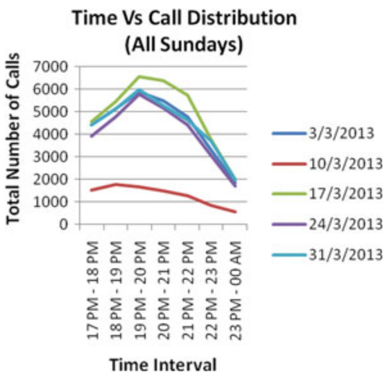


Fig. 4 Call distribution for all Sundays in May 2013

Fig. 5 Call distribution (during event time in all Sundays)



visitor count distribution. The new visitor CDRs are clustered based on tower identity. The tower-wise CDR count distribution is clearly shown in Fig. 7.

Table 1 New visitors' count using mobile CDR

New visitors' count		
Event day (intersection of incoming and outgoing CDRs)	Filtered by	
	Common user on all days (intersection)	All common users on all days (union)
3/3/2013	5704	4978
10/3/2013	454	219
17/3/2013	6318	5426
24/3/1023	5220	4354
30/3/2013	5700	4921

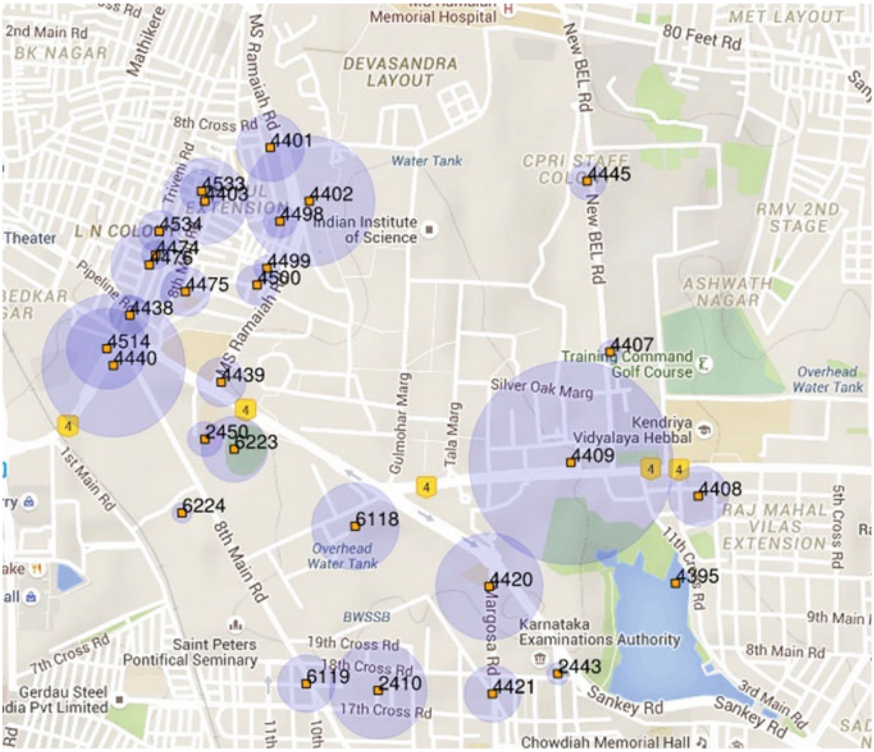


Fig. 6 Tower-wise visitor distribution

5 Conclusion

The mobile CDRs collected from single TSP are used to identify the new visitors present in an unusual event that occurred in a city. Every day the mobile CDR count is around 52 million records. The outliers, who make more number of calls in the

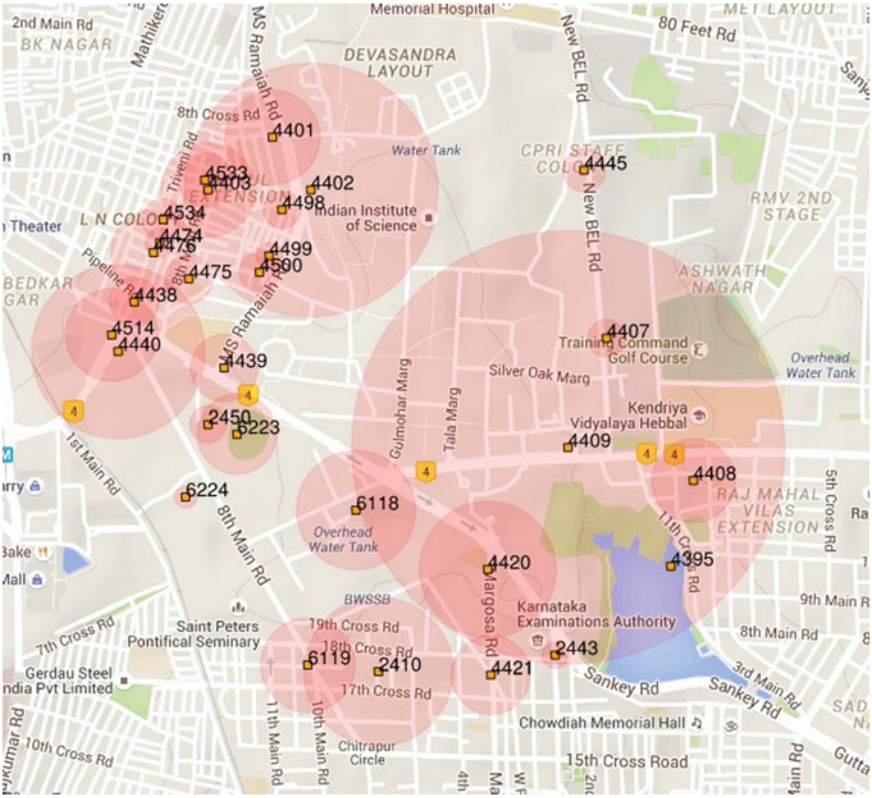


Fig. 7 Tower-wise CDRs distribution

event area, are identified and removed. New visitors present in the event are also identified and counted. The towers present in and around the event location are also identified and the CDRs are extracted from the total collected CDRs. Likewise, the duration of the visitors' stay in the event area is identified by computing MICT. The visitor count identified is less compared to all the other days. Some unusual events occurred on this day are identified. The tower-wise visitors and CDRs are analyzed and it is found that most of the visitors' associated CDRs are serviced by the towers erected near to the event location. Further research directions using CDRs are visitor level mobility between the events, visitors' behavioral analysis, population mobility changes during emergency events like earth quake and a bomb blast and classification of cities based on spatiotemporal data.

Acknowledgements I thank Indian Academy of Science (IAS) for having given me an opportunity to take up research under Summer Research Fellowship Programme at Indian Institute of Science (IISc), Bengaluru. It has provided me valuable suggestions and data set for doing the present work in Super Computer Research Centre in IISc.

References

1. Botta F, Moat HS, Preis T (2015) Quantifying crowd size with mobile phone and twitter data. R Soc Open Sci 2(5):150–162
2. Calabrese F, Di Lorenzo G, McArdle G, Pinelli F, Van Lierde E (2015) Real-time social event analytics. Netmob
3. Calabrese F, Pereira FC, Di Lorenzo G, Liu L, Ratti C (2010) The geography of taste: analyzing cell-phone mobility and social events. LNCS, pp 22–37
4. Davies L, Coleman R, Ramchandani G (2010) Measuring attendance: issues and implications for estimating the impact of free-to-view sports events. Int J Sport Mark Spons 12(1):6–18
5. Ferrari L, Mamei M, Colonna M (2014) Discovering events in the city via mobile network analysis. J Ambient Intell HumIzed Comput 5(3):265–277
6. Girardin F, Vaccari A, Gerber A, Biderman A, Ratti C (2009) Quantifying urban attractiveness from the distribution and density of digital footprints. Int J Spat Data Infrastruct Res 4:175–200
7. Hadden J, Tiwari A, Roy R, Ruta D (2007) Computer assisted customer churn management: state-of-the-art and future trends. Comput Oper Res 34(10):2902–2917
8. Lee AJT, Chen Y-A, Ip W-C (2009) Mining frequent trajectory patterns in spatial–temporal databases. Inf Sci 179(13):2218–2231
9. Louail T, Lenormand M, Ros OGC, Picornell M, Herranz R, Frias-Martinez E, Ramasco JJ, Barthelemy M (2014) From mobile phone data to the spatial structure of cities. Sci Rep 4
10. Mamei M, Colonna M (2016) Estimating attendance from cellular network data. Int J Geogr Inf Sci 1–21
11. Oliver N, Rosario B, Pentland A (1999) Graphical models for recognizing human interactions. In: Advances in neural ınformation processing systems, pp 924–930
12. Quercia D, Di Lorenzo G, Calabrese F, Ratti C (2011) Mobile phones and outdoor advertising: measurable advertising. Institute of Electrical and Electronics Engineers
13. Quercia D, Lathia N, Calabrese F, Di Lorenzo G, Crowcroft J (2010) Recommending social events from mobile phone location data. In: IEEE 2010 data mining (ICDM), University of Technology, Sydney, Australia. IEEE, pp 971–976
14. Ratti C, Frenchman D, Pulselli RM, Williams S (2006) Mobile landscapes: using location data from cell phones for urban analysis. J Environ Plan B: Plan Design 5:727–748
15. Toole JL, Lin Y-R, Muehlegger E, Shoag D, González MC, Lazer D (2015) Tracking employment shocks using mobile phone data. J R Soc Interface 12(107):2015018
16. Traag V, Browet A, Calabrese F, Morlot F (2011) Social event detection in massive mobile phone data using probabilistic location inference. In: IEEE 2011 privacy, security, risk and trust (PASSAT) and 2011 IEEE third international conference on social computing (SocialCom) 2011 IEEE third international conference, MIT Media Lab, Boston, MA, USA. IEEE, pp 625–628
17. Williams NE, Thomas TA, Dunbar M, Eagle N, Dobra A (2015) Measures of human mobility using mobile phone records enhanced with GIS data. PloS One 10(7):e0133630

Drunk Driving and Drowsiness Detection Alert System

Vivek Nair and Nadir Charniya

Abstract Advancement of safety features to avert drunk and drowsy driving has been one of the leading technical challenges in the automobile business. Especially in this modern age where people are under serious work pressure has led to higher crash rates. To prevent such accidents this paper discusses the use of nonintrusive techniques by using visual features to determine whether driver is driving in alert state. Drowsiness detection has been implemented using HAAR Cascade for face and eye closure detection and yawn detection implemented using Template matching in visual studio 2013. For drunk state detection, an alcohol sensor (MQ-3) has been implemented to avoid drunk driving. If the driver is found to be in drunk or drowsy condition, then an alarm would be generated and the driver being alerted using a buzzer and a vibrator that can be placed in the seatbelt or under driver seat thus preventing from mishaps taking place.

1 Introduction

Scientists have been working for over a decade on designing driver inattention monitoring framework. Over the years, several improvements in driver safety have been made yet a significant number of serious accidents still occur all over the world. Driver drowsiness and drunk state being the major reasons. Each year approximately 60,000 automobile accidents occur due to sleepiness related problems. Studies indicate that 25–30% of driving accidents are related to drowsiness [1]. Major approaches developed to detect driver inattention are classified as physiological, driving-behavior-based, and visual-feature-based approaches. In pragmatic applications, visual-feature-based approaches are preferred since they are naturally nonintrusive to the driver [2]. The most common usage of yawning detection is in driver

V. Nair · N. Charniya (✉)
EXTC, V.E.S. Institute of Technology, Mumbai, India
e-mail: nadir.charniya@ves.ac.in

V. Nair
e-mail: nvivek94@gmail.com

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_113

fatigue detection systems, where yawning is one factor among others, such as percentage eye closure, eye blink rate, etc. [3]. The U.S. national highway traffic safety administration (NHTSA) fatality analysis reporting system encyclopedia shows that there were approximately 55,926 vehicles involved in collisions in 2007 [4]. India revealed that about 40% of the road accidents have occurred under the influence of alcohol in a study done by alcohol and drug information centre (AIDC) [5].

2 Project Objective

This work aims towards saving valuable lives and prevent road accidents by developing a framework to avoid drunk and drowsy driving. The work has been developed using visual features that are nonintrusive in nature and can easily detect driver drowsiness which is monitored using HAAR Cascade files from OpenCV, yawn detection via template matching and drunk state using MQ-3 sensor. An alert is generated if driver is found to be drunk or drowsy state with help of buzzer and vibrator, thus, providing an all-round protection to driver and others present in the automobile.

3 Face and Eye Detection

Techniques using visual features have been implemented using computer vision approach for the detection of drowsiness. Exploiting visual features focuses on extracting facial features like face, eyes, and mouth [6]. Analyzing the state of eyes and mouth can provide observable cues for drowsiness detection process. This section focuses on how face and eyes have been successfully detected using the HAAR Cascade, which is based on Viola–Jones algorithm [7].

Working of Face detection is as described.

The working of Face detection using HAAR Cascade is as shown in Fig. 1. Where in first the image acquired on system start up, the images are gray-scaled and histogram equalization done for faster operation and better detection. The HAAR Cascade files were used from OpenCV library to detect the face which was marked with rectangle on detection. Here faces detected are marked with rectangles as shown in Fig. 2. If more than one face found by camera, then the next step was to find the largest face (as per the largest rectangular area found; as the driver, will be the one placed closest to the camera) and then set the largest face as Region of Interest (ROI). Driver's eyes were detected from the ROI as shown in Fig. 3 using HAAR Cascade for faster computation time.

Figures 3 and 4 shows the eyes being detected from the largest detected face. Figure 5 provides with the overall flow on how drowsiness of the driver would be detected using visual features. As described earlier once the face is detected, the largest face is set as ROI and from which the eye was detected using HAAR Cascade

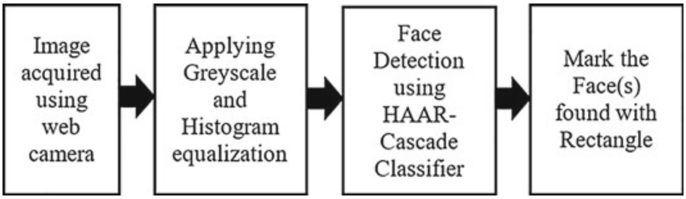


Fig. 1 Steps used to for face detection

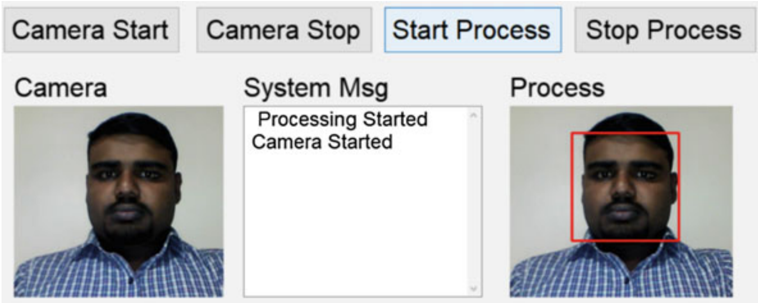


Fig. 2 Face detection using HAAR Cascade classifier

Fig. 3 Detection of eyes from face

[8]. If the eyes were found to be in closed state for a duration of three seconds or more an alert would be generated with buzzer and vibrator to alert the driver.

4 Yawn Detection

Yawn detection has been implemented using Template matching. Here the matching done between the template patch sliding over the image patch using normalized correlation coefficient. There were six methods overall to find template matching [9]. Normalized Correlation Coefficient provided with the best results as shown in Fig. 6.

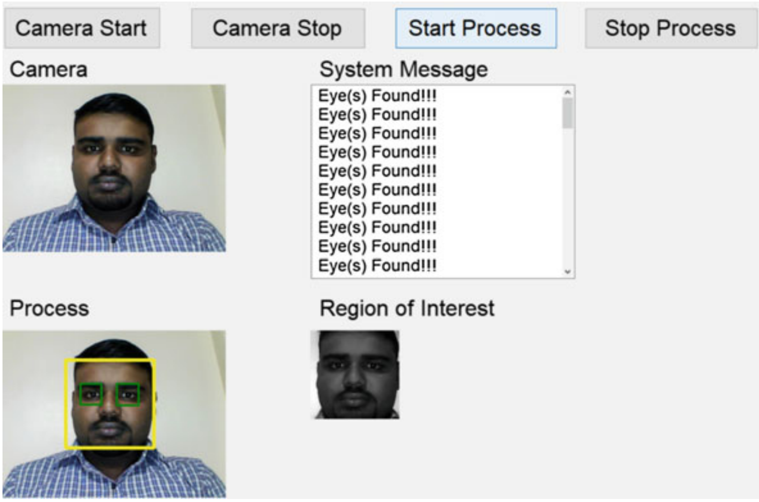


Fig. 4 Results for eye detection

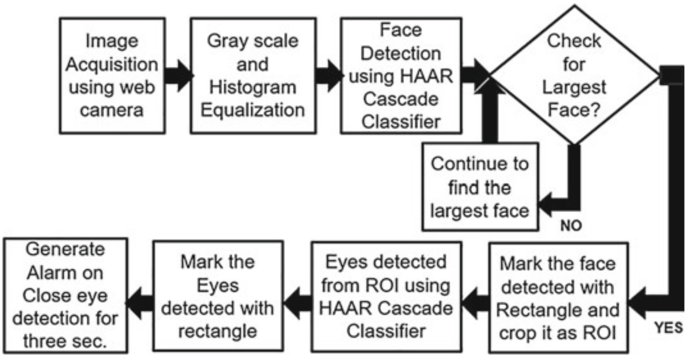


Fig. 5 Flow diagram to determine drowsiness of a driver using visual features

$$R(x, y) = \frac{\sum_{x',y'} \left(T'(x', y') \cdot I'(x + x', y + y')\right)}{\sqrt{\sum_{x',y'} T'(x', y')^2 . \sum_{x'y'} I'(x + x', y + y')^2}}, \quad (1)$$

where,

$$T'(x', y') = T(x', y') - \frac{1}{(w.h). \sum_{x'',y''} T(x''y'')}. \quad (2)$$

$$I'(x + x', y + y') = I(x + x', y + y') - \frac{1}{(w.h). \sum_{x'',y''} I(x + x'', y + y'')}. \quad (3)$$

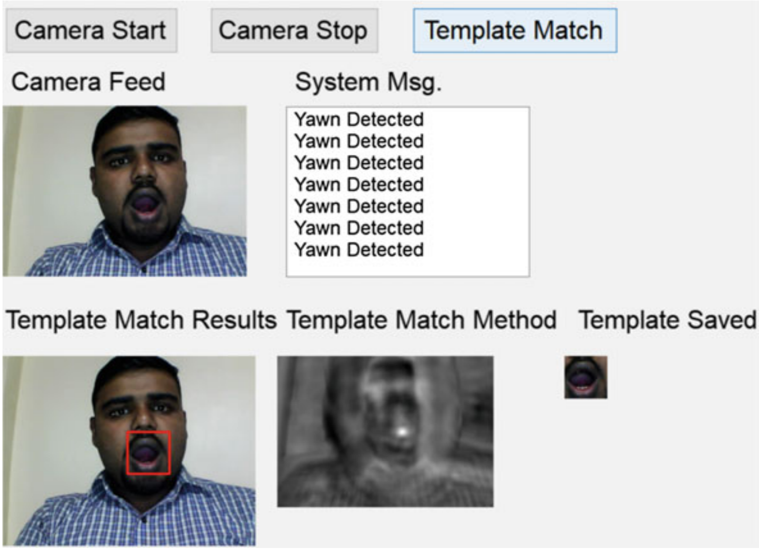


Fig. 6 Yawn detection using normalized correlation coefficient

$I_{(x,y)}$	Value of image pixel in location (x, y)
$T_{(x,y)}$	Template value of image pixel in location (x, y)
$R_{(x,y)}$	Resultant value in location (x, y) for Normalized Correlation Coefficient
Source Image	W (Width).H (Height) of Source image pixels.
Template Image	w (width).h (height) of Template image pixels.
Size of resulting images	$W - w + 1.H - h + 1$ pixels.

Yawn detection as shown in Fig. 7 has been divided into two stages: training and testing respectively. Figure 7a shows the training stage; on image being acquired, manually select the yawn template of the driver via a mouse drag over yawn region and then resize the image and save the yawn template this concludes the training part. Figure 7b describes the testing phase where we first set a threshold value for with a match would be generated and perform template matching using the correlation coefficient normalized method. On yawn detection, a rectangle will be drawn of the area where yawn has been detected and generate an alert with help of buzzer and vibrator to alert the driver.

5 Drunk State Detection

For drunk state detection, a simple MQ-3 alcohol sensor has been used which easily detects ethanol in the air [10, 11]. This sensor has high sensitivity to alcohol and

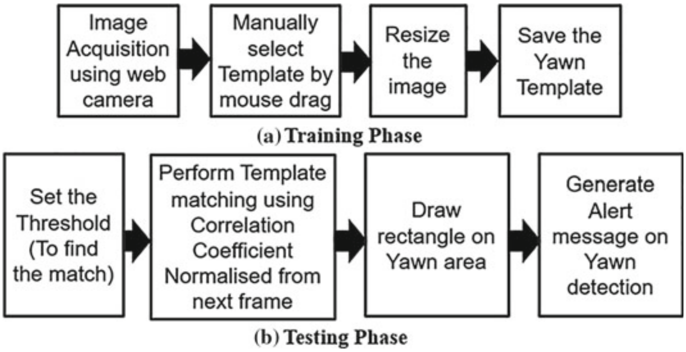
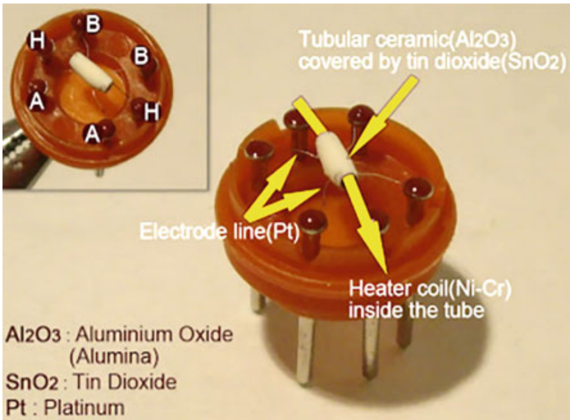


Fig. 7 Flow diagram to determine drowsiness of a driver from Yawn state

small sensitivity to Benzine. Its fast response with stable and long life makes it the best choice for drunk state detection. It is commonly used in several breath analyzers or breath testers for the detection of ethanol in the human breath. The core system is the cube which is made of an Alumina tube covered by SnO_2 (tin dioxide) as shown in Fig. 8. Coil on being heated the SnO_2 ceramics act as a semi-conductor, resulting in more movable electrons, which increases current flow. When the alcohol molecules in the air meet the electrode, ethanol burns into acetic acid resulting in more current being produced. So, greater the alcohol molecules, more current generated. This change in current results in alarm generation [12].

Figure 9 describes the working of the system. When a driver enters the car and on turning "ON" the ignition the driver first gives a breath sample to which if found to be in drunk state the emergency lights glow and the engine is turned "OFF". If driver found to be sober then the web camera is used to capture driver's visual features. A yawn template is to be provided to find yawn detection. After which the system is continuously monitoring driver's behavior. On eye closure or on yawn detection an

Fig. 8 Construction of MQ-3 Gas sensor [12]



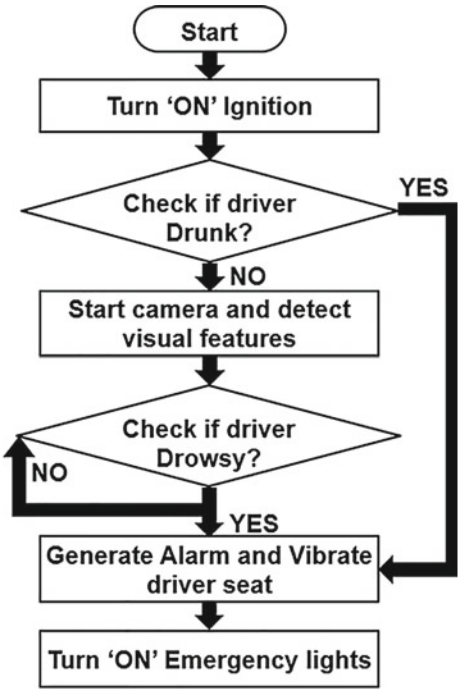


Fig. 9 The flow diagram of the entire working of system

alert is generated with help of a buzzer and vibrator motor placed (under driver seat or on seat belt) which will help the driver to get back to alert state. On the outside, the emergency lights would be turned "ON" to avoid accidents. The alert will be continuous until the driver himself turns "OFF" the alarm.

The system diagram of driver drunk and drowsy state detection is as shown in Fig. 10. An iBall C8.0 web camera with 4.0 MP video resolution at 30fps was used for capturing driver's visual features in real time which would be displayed on the PC screen. The PC is connected to the LPC2148 microcontroller via an USB to TTL connection for serial communication. Once the communication is made the system becomes ready for execution. Once the driver being found to be in drunk or drowsy state the LCD displays message stating the driver is Drunk or Drowsy respectively. The MQ-3 alcohol sensor is directly attached to the microcontroller and provides an alert on drunk state detection by turn on alarm and vibrators and turning off the engine (that is the relays are turned off) so that the driver is not allowed to drive in drunk state. The buzzer is used to generate an alarm to alert the driver and the vibrator used which can be placed under the driver's seat or at driver's seatbelt to alert the driver from drowsy state. The parking LED's or the emergency lights are used to alert the other vehicles mainly coming from the rare end of the vehicle to alert the driver of other vehicle that the vehicle in front is having some issues thus avoiding accidents from happening and providing overall safety to all passengers in the vehicle.

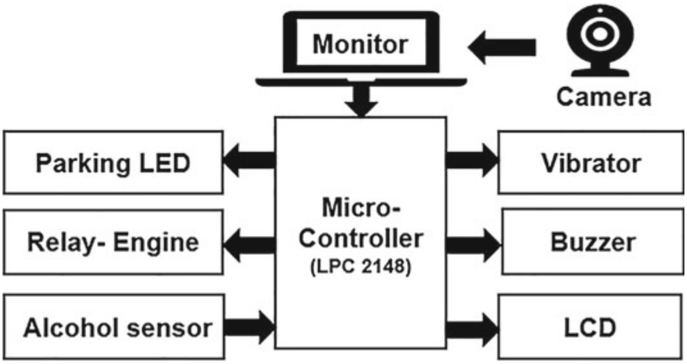


Fig. 10 System block diagram of driver drunk state and drowsiness detection with alert system

6 Results

The HAAR Cascade classifier has been implemented for face and eye detection, where in the.xml files namely haarcascade_frontalface_default and haarcascade_eye has been used to detect the face and eyes respectively in the continuous real time frame. For yawn detection template matching is used to monitor driver yawn state.

Drowsiness detection at varying lighting conditions is as follows Tables 1, 2 and 3.

Comparison between different varying conditions with 50 samples.

From the results obtained as shown in Figs. 11, 12, 13, 14, 15, 16 and 17, it is clearly visible that accuracy changes with varying lighting conditions. The best results were obtained on evenly lit facial lighting condition.

Table 1 For bright face: for 50 samples

Eye closure detected (3 s)	46
Eyes detected	46
2 eyes detected	30
False detection	04
Accuracy (%)	92

Table 2 For dark face (low light): for 50 samples

Eye closure detected (3 s)	38
Eyes detected	41
2 eyes detected	25
False detection	12
Accuracy (%)	76

Table 3 For evenly lit face: for 50 samples

Eye closure detected (3 s)	50
Eyes detected	50
2 eyes detected	42
False detection	00
Accuracy (%)	100

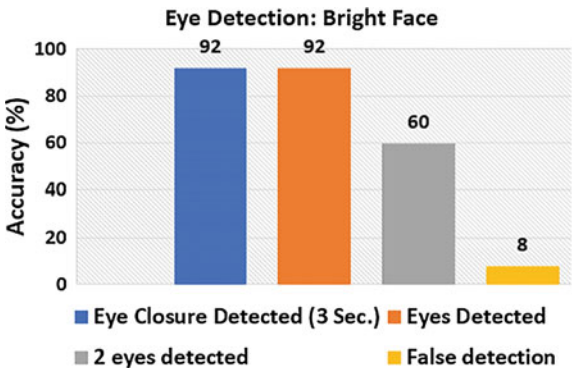


Fig. 11 Eyes detected for bright face

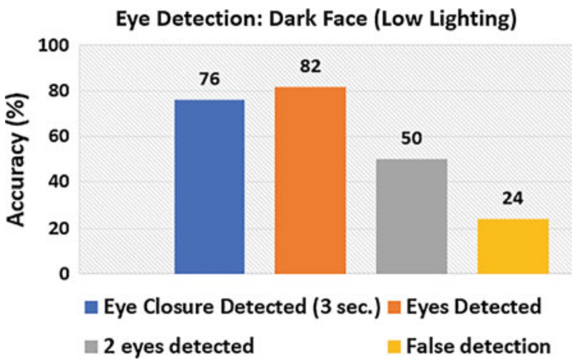


Fig. 12 Eyes detected for dark face (low lighting)

For Yawn detection: Template matching using Normalized Correlation Coefficient (C_{CN}) method has been implemented the results obtained in Tables 4, 5, 6 and 7.

From Tables 4, 5, 6 and 7 observations, it was concluded that the best threshold to be set to get the best yawn detection result for any background was 0.75.

For drunk state detection: The overall system accuracy (under ideal conditions) obtained from the best lighting condition (evenly Lit face condition) as observed in Table 3, setting a threshold of 0.75 was observed from Tables 4, 5, 6, 7 and Alcohol sensor accuracy was obtained as shown in Table 8. Table 9 shows the overall average accuracy of the system using Figs. 18, 19, 20, 21, 22, 23, 24 and 25 respectively.

Table 4 Template matching: bright background

C_{CN}	Yawn detected	Total samples	Accuracy (%)
0.95	2	20	10
0.90	14	20	70
0.85	15	20	75
0.80	19	20	95
0.75	20	20	100
0.70	20	20	100
0.65	20	20	100
0.60	19	20	95
0.55	19	20	95
0.50	0	20	0

Table 5 Template matching: black background

C_{CN}	Yawn detected	Total samples	Accuracy (%)
0.95	7	20	35
0.90	15	20	75
0.85	18	20	90
0.80	18	20	90
0.75	19	20	95
0.70	20	20	100
0.65	0	20	0
0.60	0	20	0
0.55	0	20	0
0.50	0	20	0

Table 6 Template matching: plain background

C_{CN}	Yawn detected	Total samples	Accuracy (%)
0.95	0	20	0
0.90	12	20	60
0.85	20	20	100
0.80	20	20	100
0.75	20	20	100
0.70	18	20	90
0.65	0	20	0
0.60	0	20	0
0.55	0	20	0
0.50	0	20	0

Fig. 13 Eyes detected for evenly lit condition

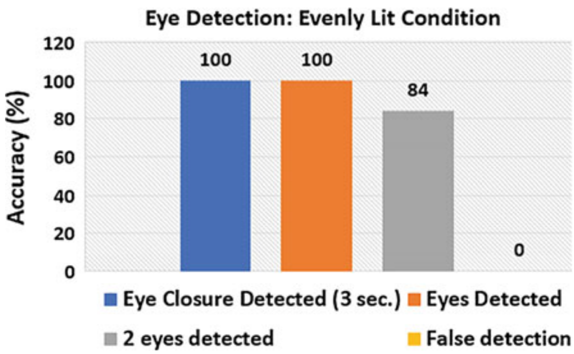


Fig. 14 Eye closure detection (3 s) for varying lighting conditions

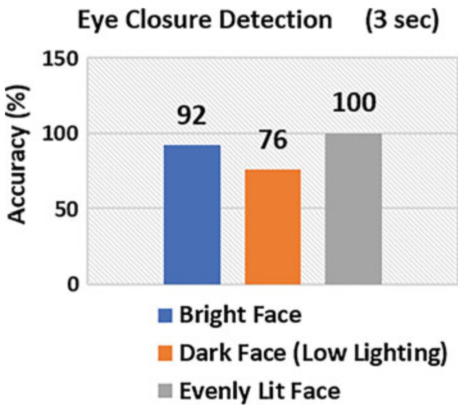


Fig. 15 Eye detected for varying lighting conditions

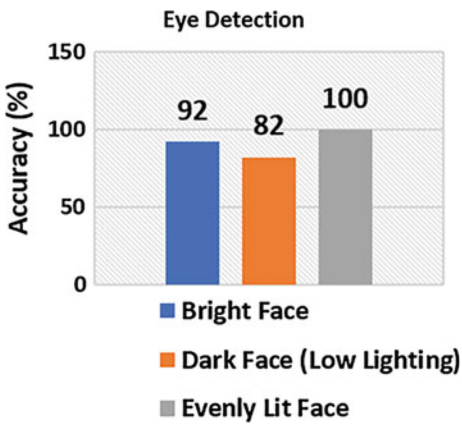
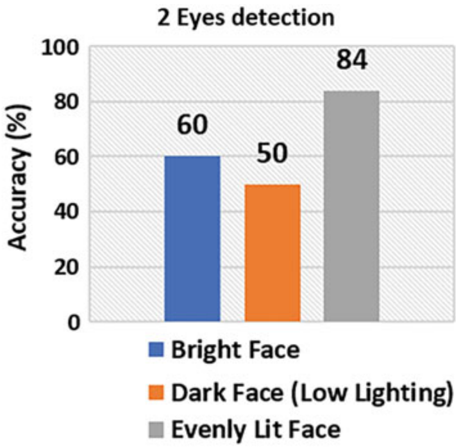


Fig. 16 Two eyes detected for varying lighting conditions



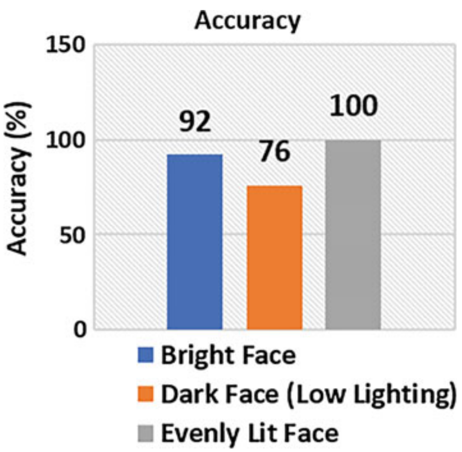


Fig. 17 Overall accuracy for varying lighting conditions

7 Conclusion

Over the decade, several drunk state and drowsiness detection techniques have been developed, even so there have been increase of accident cases due to drunk driving and driver drowsiness. This system has been implemented using nonintrusive techniques that do not bother the driver while driving thereby increasing chances to find the driver drowsy state condition and by using a simple alcohol sensor (MQ-3) to detect drunk driver condition. This framework uses the face, eye detection for drowsiness detection that has been merged with yawn detection so that in case a driver wears glasses or shades while driving, still the drowsy state can be detected using yawn results. The proposed system had very high accuracy when tested in well-lit conditions. The hardware to alert the driver during drowsy and drunk state has also been developed

Table 7 Template matching: textured background

C_{CN}	Yawn detected	Total samples	Accuracy (%)
0.95	0	20	0
0.90	2	20	10
0.85	14	20	70
0.80	19	20	95
0.75	20	20	100
0.70	20	20	100
0.65	20	20	100
0.60	12	20	60
0.55	0	20	0
0.50	0	20	0

Table 8 Alcohol sensor detection: for 30 samples

Test	Result
Detection	28
No detection	00
False detection	02
Accuracy (%)	93.33

Table 9 System accuracy for 30 samples

Readings	Results	Accuracy (%)
Eye closure detection	27	90.00
Yawn detected	29	96.67
Drunk detection	28	93.33
Overall	84	93.33

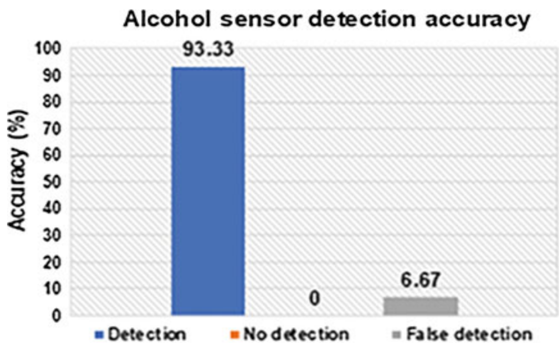


Fig. 18 Accuracy of alcohol sensor

with help of buzzer and vibrator motor along with the turning "ON" of emergency (parking) lights to alert other drivers. Here the accuracy of the entire system has

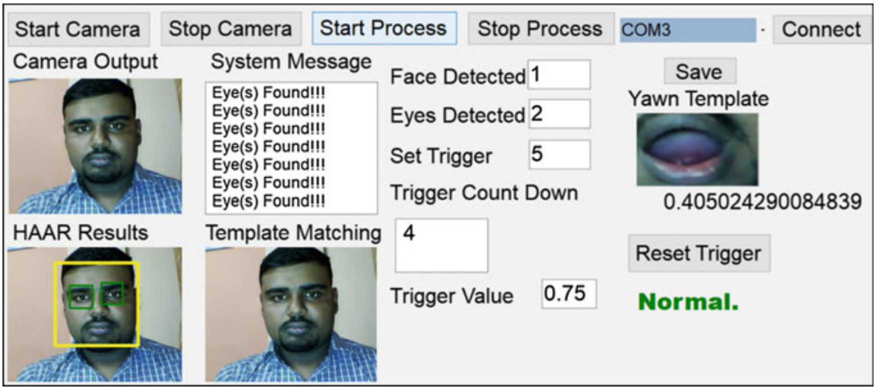


Fig. 19 Setup when driving in alert state

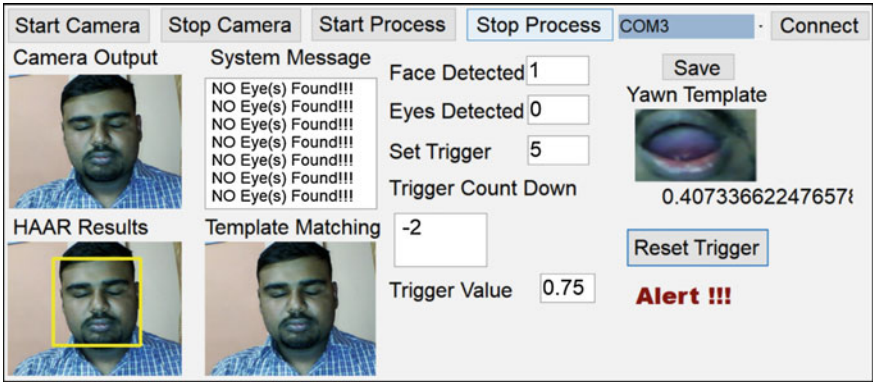


Fig. 20 Setup when driving in drowsy state (eye closure)

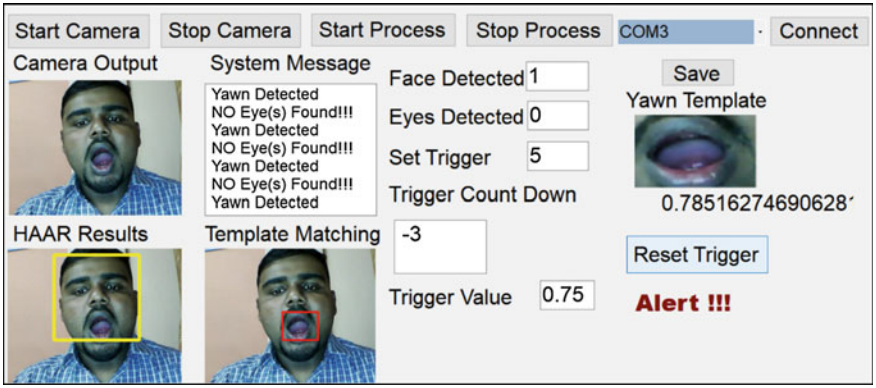


Fig. 21 Setup when driving in drowsy state (yawn detection)

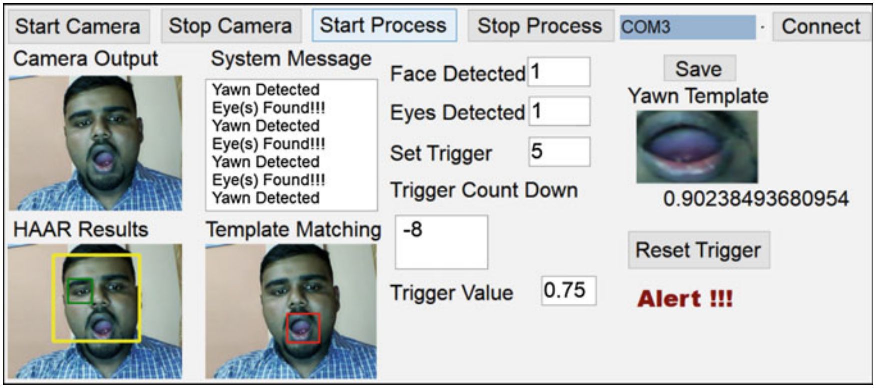
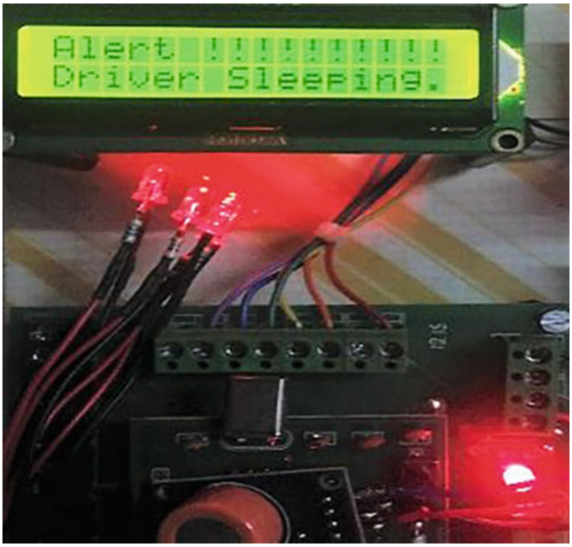


Fig. 22 Alert generation on any one condition being detection

Fig. 23 Alert message displayed on LCD on drowsiness detection



been calculated using varying lighting conditions for eye closure detection using HAAR Cascade classifier and yawn detection using template matching in different background conditions with varying threshold levels. From the observations, it was observed that the best detection was found on evenly lit conditions for eye closure detection. In addition to this, the best match for yawn detection was observed keeping a threshold of 0.75. Overall average accuracy of the system was found out to be 93.33%. The accuracy of the system could be increased under low lighting conditions using a better resolution camera.

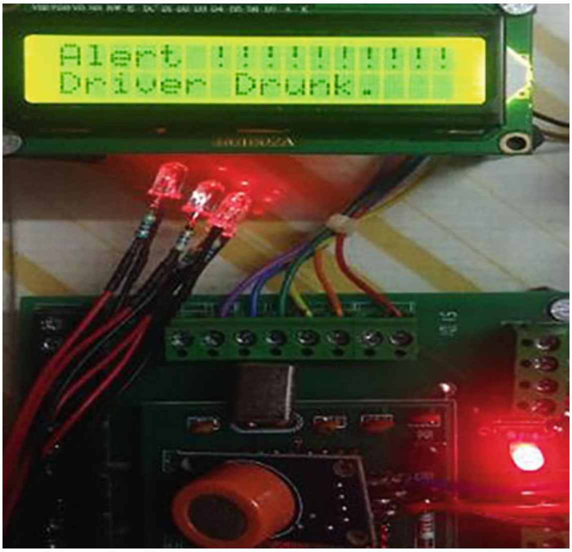


Fig. 24 Alert message displayed on LCD on drunk state detection

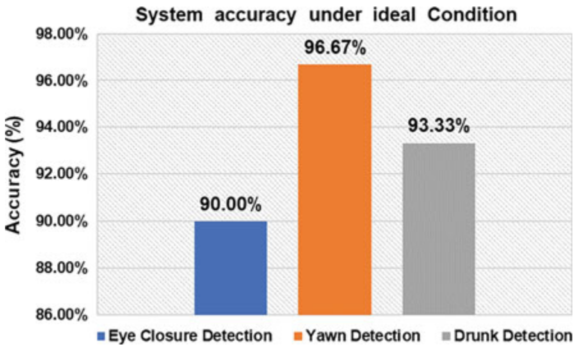


Fig. 25 Overall system accuracy of the work

Acknowledgements The authors would like to thank the anonymous reviewers for their valuable comments and suggestions. I (Vivek Nair—Author) would like to clarify that the participant (driver) is myself and give consent to be used in the study.

References

1. Kaplan S, Guvensan MA, Yavuz AG, Karalurt Y (2015) Driver behavior analysis for safe driving: a survey. IEEE Trans Intell Transp Syst 16(6):3017–3032
2. Mbouna RO, Kong SG, Chun M-G (2013) Visual analysis of eye state and head pose for driver alertness monitoring. IEEE Trans Intell Transp Syst 14(3):1462–1469
3. Omidyeganeh M, Shirmohammadi S, Abtahi S, Khurshid A, Farhan M, Scharcanski J, Hariri B, Laroche D, Martel L (2016) Yawning detection using embedded smart cameras. IEEE Trans Instrum Meas 65(3):570–582

4. Chacon-Murguia MI, Prieto-Resendiz C (2015) Detecting driver drowsiness: a survey of system designs and technology. IEEE Consum Electron Mag 4(4):107–119
5. Sivakumar T, Krishnaraj R (2012) Road traffic accidents (Rtas) due to drunken driving in India challenges in prevention. IRACST—Int J Res Manag Technol (IJRMT) 2(4):401–406
6. Bhandari GM, Durge A, Bidwai A, Aware U (2014) Yawning analysis for driver drowsiness detection. IJRET: Int J Res Eng Technol 3(2):502–505
7. Viola P, Jones M (2001) Rapid object detection using a boosted cascade of simple features. In: Conference on computer vision and pattern recognition
8. Nair VR, Charniya NN (2017) Drunk driving and drowsiness detection. In: International conference on intelligent computing and control (I2C2), Coimbatore, India, vol 2, pp 30–35
9. Kaehler A, Bradski G. Learning OpenCV3, Computer vision in C++ with OpenCV library
10. Dong W, Cheng CQ, Kai L, Bao-hua F (2011) The automatic control system of anti-drunk-driving. In: International conference on electronics, communication and control (ICECC), Ningbo, China. IEEE, pp 523–526
11. Wu Y-C, Xia Y-Q, Xie P, Ji X-W (2009) The design of an automotive anti-drunk driving system to guarantee the uniqueness of driver. In: International conference on information engineering and computer science (ICIECS), Wuhan, China. IEEE
12. Eun Jung Park (2008) Sensor report—MQ-3 Gas sensor. http://sensorworkshop.blogspot.in/2008/04/sensor-report-mq3-gas-sensor.html

A Survey on Intelligent Face Recognition System

Riddhi Sarsavadia and Usha Patel

Abstract Face recognition system is a computer's capability which gives it a vision of performing two fundamental operations the detection and the recognition of a human face. With the advancement of machine learning algorithm and image processing techniques the accuracy of face recognition system has been significantly improved. The objective of this paper is to give a detailed survey of a few face recognition algorithm with their features and limitations. The basics of face detection and face recognition techniques along with their approaches are described in the section.

1 Introduction

The human's face is a complex because of facial components which can change after some time. Humans have a very good capacity to recognize several faces learned throughout their lifetime and can identify faces in milliseconds, even after years of separation, but the system does not have the same capacity. So, for this purpose, we need to create a face recognition system to make our system as intelligent as humans. Face recognition by systems can be significant and has a huge amount of usage for security purpose for organizations, access control for higher level authorities, national defences, etc. [1]. Face recognition incorporates fundamentally three-errand face detection, feature extraction and recognition [1].

As shown in Fig. 1 the first step of the method is face identification has its own challenges which are posture invariant, the existence or deficiency of occlusion, physical components, image orientation, facial expression, and imaging conditions [2]. After detection of a face from an image, the next step would be feature extraction in which is used to extract only features that would be used for mapping with

R. Sarsavadia (✉) · U. Patel
Nirma University, Ahmedabad 382481, India
e-mail: riddhisarsavadia9@gmail.com

U. Patel
e-mail: ushapatel@nirmauni.ac.in

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_114

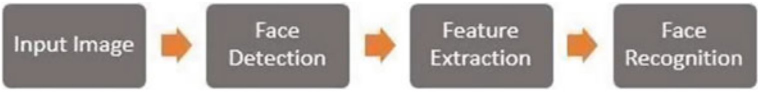


Fig. 1 Face recognition system

image database for face recognition. For face recognition step we have mainly three approaches which are the holistic approach, feature-based and hybrid approach [1].

Using Face recognition approaches, we can get recognition rate of frontal face, oriented face (up to 30°) and props (beard and eyeglasses with an existing database) up to 80–85%. To overcome these challenges, we use hybrid approaches for face recognition. When we recognize face with eyeglasses using face recognition techniques, there will be some problems due to the thickness of eyeglasses, the size of a frame of eyeglasses, reflective property of glasses etc. So, we use eyeglass removal [1] to remove eyeglasses after detecting a face. In this project work, the whole implementation would be done using OpenCV (Open Computer Vision) [3].

OpenCV (Open source Computer Vision) is a set of libraries with programming functions which has the main objective of real-time image processing [4]. OpenCV was developed by Intel's research centre by Willow Garage in 1999 at Nizhny Novgorod (Russia) [3]. It is freely available for commercial as well as non-commercial usage. The method is very helpful for developing highly efficient computer vision applications.

2 Face Detection Approaches

Some of the main face detection methods are discussed here.

(1) Some comprehensive approaches have been derived by some well-known researchers of the human faces. Their rules are robust but the fundamental hindrance in applying those ideas is the difficulty of translating human intelligence into set guidelines.
(2) Featured-based methods: Some invariable parameters of human faces are helpful in detecting face-cut, skin tone, skin texture, etc. But capabilities of determining those parameters might be severely impacted due to colour reproduction, noise, grains, etc.
(3) Template matching: Thanks to scale, pose, and shape of the human face which is given as an input, have to be matched with predefined template in the machine.
(4) Appearance-based method: In this type, the methods are learnt from the examples and references in the images. Machine learning techniques, statistical analysis approach can be applied to find relevant parameters of face from the overall image. These metrics will be used in overall analysis of the input. However, in the template matching, the templates and methods are predefined by the researchers.

A part of the primary face detection techniques is mentioned below.

(1) Defining the right face matching techniques are defined by the rules and regulations provided by the experts. The main challenge in this method is to analyse the given image and comprehend it as the human element, must be done in very limited set of rules.
(2) Featured-based techniques: Invariable aspects of appearances are monetize for identifying surface, skin shading. Be that as it may, the output from such machine calculation can be extremely underestimated because of light, focus, exposure, etc.
(3) Template coordinating: The input of this method is manipulated in the confined template. Be that as it may, the execution here suffers because of the undefined parameters like shape, scale and posture.
(4) Appearance-based technique: In these types of strategies, the agendas are predefined by experts. While, the formats in the other strategies are found from the cases in images. Machine Learning and statistical investigation are the procedures which can be manipulated to find the significant qualities of face and other.

The main purpose of face discovery is to reveal that if there are any faces in the image or not and if face is present then return an extent of each face. For face detection step of our IFRS, we need to detect face using some facial components like eyes, nose, mouth, face contour with some facial features like with beard or without beard, with eyeglasses or without eyeglasses and with an orientation of face. There are several techniques which includes these cases which are shown in Table 1.

As shown in Table 1, Haar classifier can exist for multiple face detection with facial features like with eyeglasses and without eyeglasses, with beards and without beards can be detected as well as it is faster among previous approaches and we can detect faces with more than 45° orientation of faces and higher detection rate among all approaches discussed earlier.

3 Feature Extraction

The previous unit gives discussion about face detection, now, in this area, well discuss how to remove valuable and minimal highlights for further recognition procedure. Figure 2 shows basic flow of feature extraction after face detection.

When input image to an algorithm is too large for the further processing then this step is necessary. Using feature extraction, we can get reduced size of an image for further procedure by which we can get lesser time for whole procedure as compared to without reducing an image. Most of the methods for feature selection and extraction are also used for later steps. Some of them are described in Table 2.

Table 1 Comparative analysis of face detection approaches

Papers	Technique	Components	Working cases	Issues
"Human face detection in a complex background", G. Yang and T. S. Huang	Hierarchical knowledge based [2]	Eyes, nose, mouth	Frontal faces in uncluttered scenes	Translating knowledge into rules, detect faces in different poses
"Finding faces in cluttered scenes using random graph matching" T. Leung, M. Burl, and P. Perona	A computational approach to edge detection [7]	Edge of face, nose, eyes	Edges detected closed to true edges	Real world conditions and scaling and thresholding of image
"An introduction to face recognition technology" S.-H. Lin	Integration of skin colour, size and shape [1]	Skin colour, edge of face, size	Detect face at different orientations up to 45° and for facial features (glasses and beards also)	Lightning conditions, only single face
"Automatic face identification system using flexible appearance models" A. Lanitis, C. J. Taylor, and T. F. Cootes	Deformable templates: active shape model (ASM) [8]	Edge of face, eyes, eyebrows, nose, mouth, lips	Single frontal face	Multi-faces, variation in pose, scale and shape
"Eigenface-based facial recognition" D. Pissarenko	Eigenfaces for recognition [9]	Eyes, nose, mouth, face contour	Only frontal faces	Lighting conditions, facial expression, orientation
"Human face detection in a complex background" G. Yang and T. S. Huang	Example based learning for view-based human face detection [2]	Eyes, nose, mouth, eyebrows, face contour, lips	Can detect faces when lightening is nearly frontal and less computational time than eigenface	Lightning conditions
"Detecting faces in images: a survey" M. H. Yang, D. J. Kriegman, and N. Ahuja	Neural network based face detection [10]	Eyes, nose, mouth, face contour, lips	Capture complex class conditional density	Requires external tuning
"Rapid object detection using a boosted cascade of simple features" P. Viola and M. Jones	Haar classifier: robust real-time face detection by Viola and Jones [11]	Eyes, nose, mouth, face contour, lips	Detect face faster than 15 times of other approaches with different facial conditions	Requires more memory

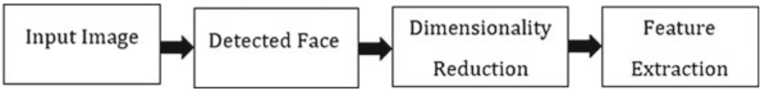


Fig. 2 Flow of feature extraction

Table 2 Comparative analysis of feature extraction

Technique	Working cases	Issues
Principle component analysis (PCA), Kernel PCA, weighted PCA [12]	Eigenfaces and dimensionality reduction	The transformed dimensions will be uncorrelated from each other, only uses second order statistics
Linear discriminant analysis (LDA), Kernel LDA [13]	Eigenvector based with supervised map; LDA with kernel method	Only small sample size problem can be solved
Independent component analysis (ICA) [14]	Transformed dimensions are independent and higher order statistics are used for dimensionality reduction	Cannot linearly separate independent sources in smaller subspace
Linear binary patterns (LBP) [4]	Uniform binary patterns will contain at most two-bitwise transactions from 0 to 1 or vice versa	Image may be blurred

4 Face Recognition Approaches

As discussed in the previous section, the images which are reduced using feature extraction technique would be used for recognition process to decrease the comparison time of an image with the database. There are various techniques for FRS [5] which can recognize human face within lesser time, using facial features like with eyeglasses or without eyeglasses, with bread or without bread, multiple face recognition at a time, orientation of face etc. [6]. IFRS techniques which incorporate these characteristics are discussed in Table 3.

5 Survey Analysis

There are various approaches for face recognition which are discussed in the previous section. In this section, complete comparative analysis of face recognition techniques with respect to real-time approach along with some facial components like facial features, props, orientation, lightning conditions, etc. are described here, as shown in Table 4, combination of different algorithms and techniques gives the accurate result, i.e. Hybrid techniques for real-time implementation of the system.

Table 3 Comparative analysis of face recognition approaches

Approach	Technique	Working cases	Issues
Holistic	Principal component analysis [12]	Requires full facial information	Facial features like with eyeglasses and with bread etc. and multiple faces at a time
Holistic	Two dimensional PCA: a new approach to appearance-based face representation and recognition [15]	Recognize single face with up to 45° orientation	Multiple faces
Holistic	Face recognition using LDA-based algorithms [13]	Recognize more than 45° oriented face	Recognize maximum 5 faces at a time
SVM	SVM: training support vector machines: an application to face recognition [16]	SVM classifies facial features using a hyper-plane	Lightning conditions and only frontal face can be recognized
Feature based	Face recognition by elastic bunch graph matching (EBGM) [17]	Faces as a graph	Recognize up to 22° oriented faces
Feature based	Face recognition by active appearance models [18]	Facial shapes, facial expressions etc.	Lightning conditions

Table 4 Survey analysis

Components	Algorithms					Hybrid techniques	
	PCA	ICA	LDA	LBP	EBGM	Fisherfaces	Proposed sys
Features							
Eyes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Nose	–	–	–	–	Yes	–	–
Ears	–	–	–	–	Yes	–	–
Face contour	–	–	–	–	–	–	–
Mouth	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Props							
Eyeglasses							
With existing dataset	Yes	Yes	Yes	–	–	Yes	Yes
Without existing dataset	–	–	–	–	–	–	Yes
Beards	–	–	–	–	–	–	–
Orientation							
Up to 45°	–	–	–	–	–	–	Yes

6 Conclusion

The major approaches for face recognition are discussed in this survey paper. The complete comparative survey analysis of face recognition techniques with respect to real-time approach shows that implementing hybrid techniques, i.e. combination of various techniques gives better and accurate results than implementing single technique for building intelligent face recognition system.

References

1. Lin S-H (2000) An introduction to face recognition technology. Inf Sci 3(1):1–8
2. Yang G, Huang TS (1994) Human face detection in a complex background. Pattern Recogn 27(1):53–63
3. https://en.wikipedia.org/wiki/Computervision, “Computer vision.”
4. https://www.learnopencv.com, “Opencv.”
5. Jain AK, Li SZ (2011) Handbook of face recognition. Springer
6. Craw I, Tock D, Bennett A (1992) Finding face features. In: European conference on computer vision, pp 92– 96. Springer
7. Leung T, Burl M, Perona P (1995) Finding faces in cluttered scenes using random graph matching. In: Proceeding of IEEE PAMI conference, pp 637–644
8. Lanitis A, Taylor CJ, Cootes TF (1995) Automatic face identification system using flexible appearance models. Image Vis Comput 13(5):393–401
9. Pissarenko D (2002) Eigenface-based facial recognition, vol 1, pp 4–9
10. Yang M-H, Kriegman DJ, Ahuja N (2002) Detecting faces in images: a survey. IEEE Trans Pattern Anal Mach Intell 24(1):34–58
11. Viola P, Jones M (2001) Rapid object detection using a boosted cascade of simple features. In: Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition, 2001. CVPR 2001, vol 1. IEEE, pp I–511
12. Moon H, Phillips PJ (2001) Computational and performance aspects of pca-based face-recognition algorithms. Perception 30(3):303–321
13. Lu J, Plataniotis KN, Venetsanopoulos AN (2003) Face recognition using lda based algorithms. IEEE Trans Neural Netw 14(1):195–200
14. Bartlett MS, Movellan JR, Sejnowski TJ (2002) Face recognition by independent component analysis. IEEE Trans Neural Netw 13(6):1450–1464
15. Yang J, Zhang D, Frangi AF, Yang J-Y (2004) Two-dimensional pca: a new approach to appearance-based face representation and recognition. IEEE Trans Pattern Anal Mach Intell 26(1):131–137
16. Hearst MA, Dumais ST, Osuna E, Platt J, Scholkopf B (1998) Support vector machines. IEEE Intell Syst Appl 13(4):18–28
17. Wiskott L, Fellous J-M, Kuiger N, Von Der Malsburg C (1997) Face recognition by elastic bunch graph matching. IEEE Trans Pattern Anal Mach Intell 19(7):775–779
18. Edwards GJ, Cootes TF, Taylor CJ (1998) Face recognition using active appearance models. In: European conference on computer vision. Springer, pp 581–595

Real-Time Health Monitoring System Implemented on a Bicycle

Rohith S. Prabhu, O. P. Neeraj Vasudev, V. Nandu, J. Lokesh and J. Anudev

Abstract It is evident that many systems have been developed and are being developed these days related to health monitoring. The main requirement is the continuous real-time health monitoring where the data, i.e. the body vitals/health parameters, can be easily understood by the user through an application interface, and this would also be shared with the corresponding physician who can be aware of the patient's vitals at all times. This is a study of methods where a real-time health monitoring system can become an integral part of the society whereby its intention mainly aimed at gaining awareness of each ones health and its implementation leading to focus on a health record which is linked to web also information reaching the doctors in time for continuous monitoring. The obese people are able to control their weight. The project being 'Real Time Health Monitoring System Implemented On A Bicycle' aimed at designing and assembly of a non-invasive health monitoring system. The conversion of cycle energy will also be taken into account to supply for the health monitoring devices and charging of the display.

R. S. Prabhu · O. P. Neeraj Vasudev · V. Nandu · J. Lokesh · J. Anudev (✉)
Department of Electrical and Electronics Engineering, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Amritapuri, India
e-mail: anudevj@am.amrita.edu

R. S. Prabhu
e-mail: rohithsprabhu96@gmail.com

O. P. Neeraj Vasudev
e-mail: neerajpaderi@gmail.com

V. Nandu
e-mail: nandhuvijay009@gmail.com

J. Lokesh
e-mail: lokeshkj006@gmail.com

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_115

1 Introduction

- At least 200,000 deaths from heart disease and strokes every year could be prevented by monitoring vital signals regularly [1, 2]. Nowadays, real-time health monitoring systems are required not only for patients to monitor their own health parameters but also for their doctors to be able to continuously assess the patient [3].
- Relating to the system presented here the patient's heart rate, calories burned, blood pressure and body temperature are mainly observed. A methodology where doctors can have full supervision and be able to understand the daily routine will help in better treatment of the patient.
- People can manage their daily routine check-up at home. In addition, this is important to provide people with continuous monitoring in non-clinical environments. However, such health management only can be achieved if the computer-based portable monitoring devices with smart sensor technologies are available [4].

2 System Architecture

See Fig. 1.

2.1 Health Monitoring

The real-time patient health monitoring system is one which incorporates various physical parameters of a human body like heart rate, blood pressure and body tem-

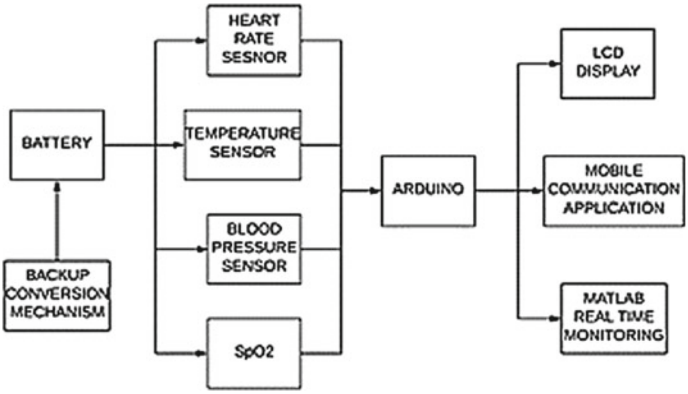


Fig. 1 Overview block diagram of the system

perature [5, 6]. This model is set to be implemented to acquire the pulse signal received from our body and is further used to measure heartbeat of a patient. Calories burned will be displayed in addition to this analysis of the ECG waveform is also considered. A microcontroller board can be used for analysing inputs from the patient. The display will be given to the mobile application, LCD display for the miscellaneous readings and continuous monitoring of the user carried out by the doctor.

2.2 Energy Conversion

The conversion part is considered to give the supply for the displays, phone and reuse of energy is seen. The dynamo set-up is used to convert the kinetic energy to electrical which after getting sufficient energy for recharging the rechargeable battery. We can step down to the required voltages of the system in which the supply voltage is required. Charging of the phone is also evident in this system.

3 Health Parameters

Here below are the methods proposed for the calculation of heart rate, temperature, blood pressure and calories. The following seem to be easy to incorporate into the system and ability to get the desired routine output is considered.

3.1 Heart Rate

This is a module designed to easily connect to the Arduino to be able to harness the heart rate of the person. It poses to be a very efficient design as the implementation and assembly onto the bicycle becomes easy. It is seen that the beats per minute are displayed in the LCD display hereby it shows the value that is seen in the serial monitor of the Arduino where usually from this which will be taken as the input directly into the system for the patient monitoring (Figs. 2 and 3).

The whole system integrated together as shown in Fig. 4 brings about heart rate monitoring system. The alert messaging, cloud storing is illustrated, and the possibility for the real-time monitoring considered makes this system one of the useful even in remote areas as the patients can have up to data assessment from their physicians.

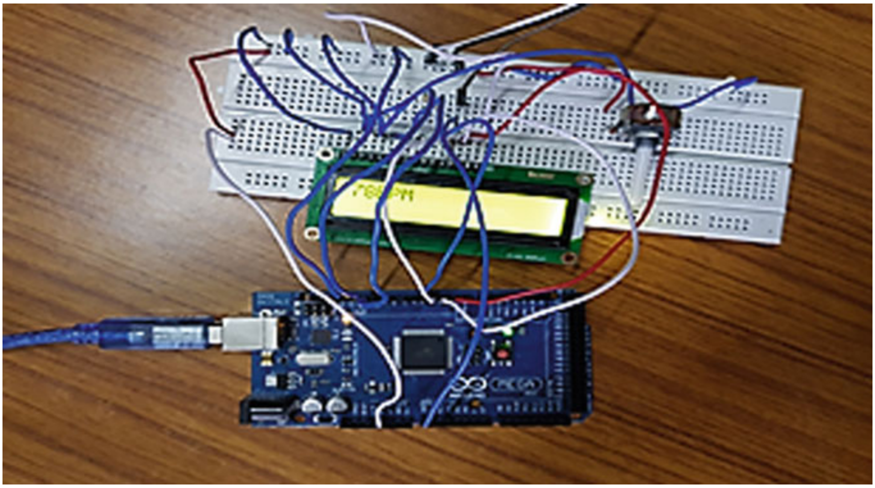


Fig. 2 Prototype displaying BPM

3.2 Calories

In this system, the method where input is heart rate monitored and are theoretically able to get the calories burnt of the person who will be using the system. The required contents needed from the formulae will be given as input by the user.

$$\text{Male:}\ \frac{((-55.0969) + (0.6309 \times \text{HR}) + (0.1988 \times W) + (0.2017 \times A))}{(4.184) \times 60 \times T} \quad (1)$$

$$\text{Female:}\ \frac{((-20.4022) + (0.4472 \times \text{HR}) - (0.1263 \times W) + (0.074 \times A))}{(4.184) \times 60 \times T} \quad (2)$$

where

HR Heart rate (in beats/minute)
W Weight (in kilograms)
A Age (in years)
T Exercise duration time (in hours)

With this, it is great for obese patients as they can set the required target and be able to achieve it with the supervision of their doctors.

The heart rate is efficiently taken in a short time and with less expense without using time-consuming and expensive clinical pulse detection systems then calculates the calories burned during an exercise from measured heartbeat [7].

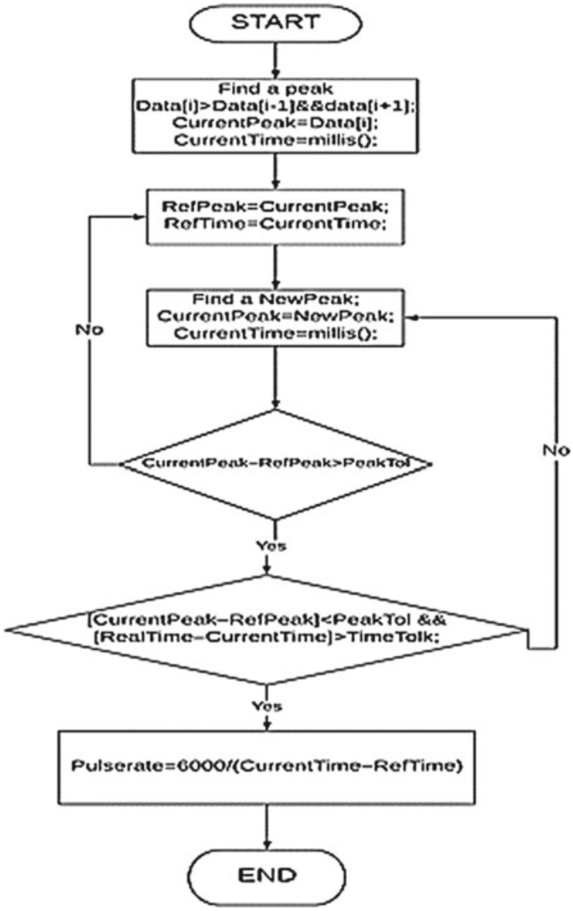


Fig. 3 Flowchart of the algorithm for getting BPM

3.3 Blood Pressure

The blood pressure is considered as the good indicator of the status of cardiovascular system. Doctor's prefer sphygmomanometer, being the most accurate device, to measure the blood pressure presently. A sphygmomanometer consists of occlusive cuff which is wrapped around the arm and a column of mercury to find the systolic and diastolic pressure. When the doctor exerts pressure on the occlusive cuff, an audible sound is produced and this sound can be heard by a stethoscope. The audible sound is known as 'korotkoff'.

The readings from the sphygmomanometer at a point when korotkoff sound start is a measure of systolic pressure and at a point where korotkoff sound stops on

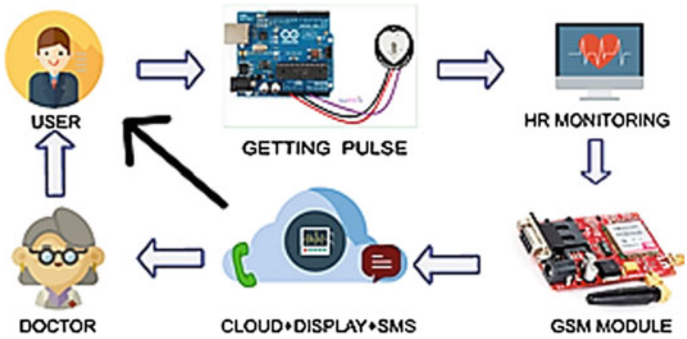


Fig. 4 Visualization of the heart rate monitoring system

lowering occlusive pressure is a measure of diastolic pressure. The blood pressure is represented in mm hg [8].

Flowchart shown in Fig. 5 gives a better understanding of how blood pressure is measured.

The pressure occurred in the arteries when the ventricles are squeezing out blood under high pressure is known as systolic blood pressure and the pressure occurred when the ventricles are filling up with the blood is known as diastolic blood pressure. This diastolic blood pressure will be low pressure compared to systolic. The net amount of blood pump out from the heart with respect to time will be the cardiac output of a person and the net amount of blood returned to the heart with respect to time is termed as venous return. For a normal healthy person, the cardiac output is equal to the venous return; this will be approximately equal to 5 L/min. The whole system integrated together as shown in Fig. 6 brings about BP monitoring system. The physician will have full control over the workout and health variations in each scenario as the block diagram below represents [9].

3.4 Temperature

LM35 is a temperature sensor which can measure the temperature more accurately compared to a thermistor. The sensor circuit is sealed and not subjected to oxidation. LM35 generates a higher output voltage than thermocouples and may not require the output voltage to be amplified. LM35 output voltage is proportional to Celsius temperature. Its scale factor is 0.1 V/°C. It does not require any external calibration or trimming and it maintains an accuracy of ±0.4 °C at room temperature and ±0.8 °C for range of 0–100 °C. It has low self-heating capability. LM35 Circuit Diagram draws only 60 μA from its supply [10]. Basically, it is operated under 4–30 V. Differential amplifiers are used. There are two inputs like non-inverting (+) and inverting (−) for the temperature sensor and an output pin (Fig. 7).

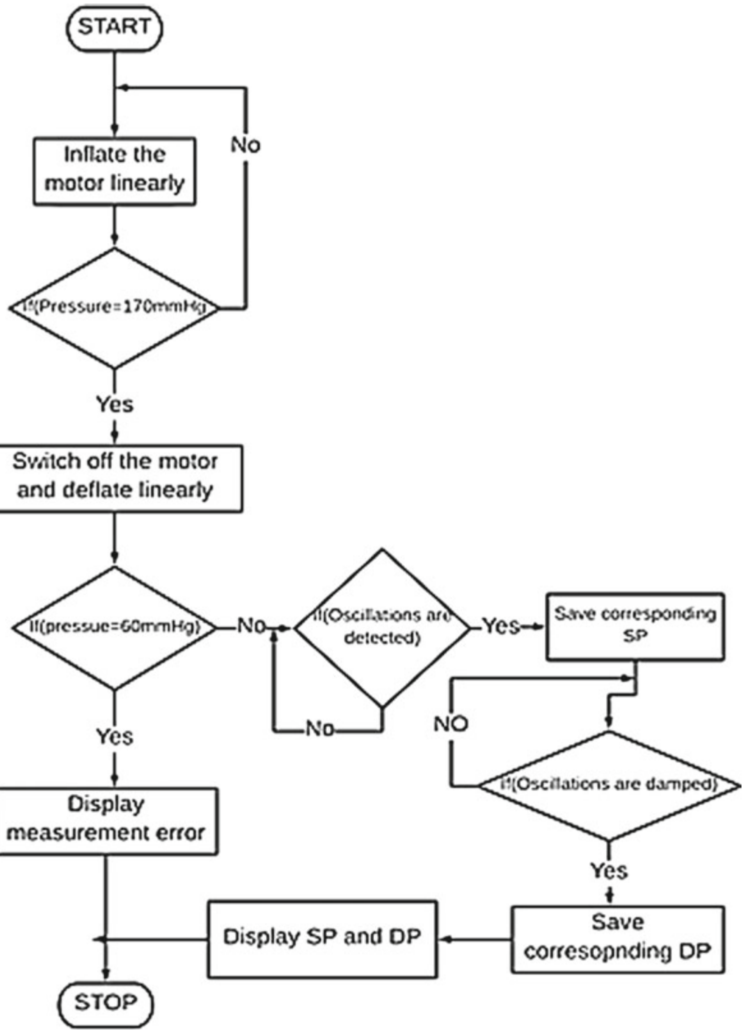


Fig. 5 Flowchart of the blood pressure monitoring system

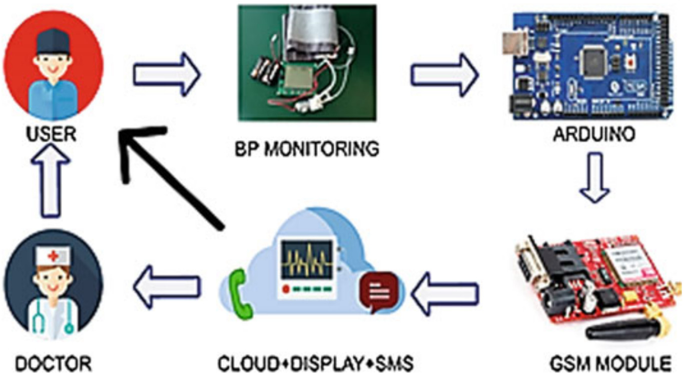
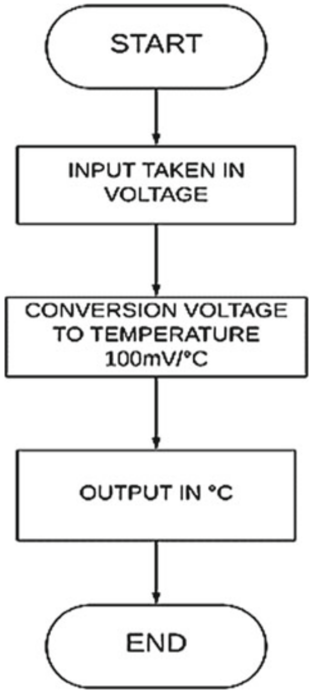


Fig. 6 Visualization of the blood pressure monitoring system

Fig. 7 Flowchart of temperature monitoring system



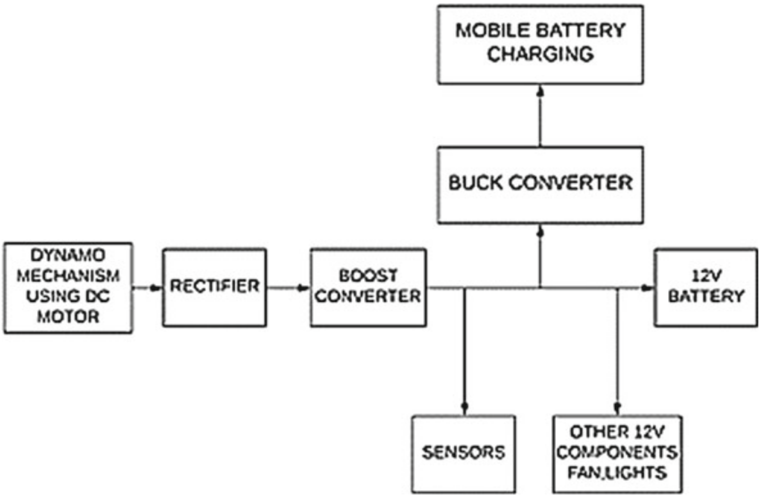


Fig. 8 Energy conversion block diagram

4 Bicycle Conversion

The conversion part plays an integral part where all the charging of the displays and system supply is provided by regeneration using the dynamo technique (Fig. 8).

4.1 Cycle Dynamo

Cycle dynamo is an energy harvester that can convert the mechanical energy produced while riding bicycles into electrical energy and thus can be utilized to power electronic gadgets requiring low power. Dynamos are available in many ranges but most of these are rated such that to deliver up to 1 A of current. This value of current is insufficient to charge the 12 V battery of the proposed system. Hence, an alternative mechanism of current generation has been implemented using a DC geared motor. The rating of the motor is such that it is having a nominal voltage of 12 V at 1000 RPM. The full load current can go up to 7 A. A current of 2 A can be delivered if the output power of the motor is 25 W. A wheel will be fixed on the shaft of the motor and is forced to touch the cycle wheel using spring mechanism. The radius of the wheel attached to the shaft can be determined using these calculations:

Normally, cycle wheel diameter varies from 45 to 60 cm.
So we take $D = 45$ cm (minimum value).
So circumference of the wheel $= [pi * D] = 140$ cm [approx.].
Assume that the cycle rotates at 300 rpm (particular case).
Let the rpm of motor be 1000 rpm.

So the ratio of rpm of cycle wheel to the rpm of the wheel attached to the shaft = 1:3 (approx.).
So that the circumference of the wheel attached to the shaft = 140/3 = 46.66 cm.
So radius = 46.66/(2 * *pi*) = 7.43 cm.
Hence, we can use motor shaft wheel of radius greater than 7 cm for this system.

4.2 Rectifier

Depending upon the polarity of rotation of the shaft, output from the dynamo can be varying. To achieve a steady and smooth direct current output, a full wave rectifier circuit is employed.

4.3 Converters

4.3.1 Boost Converter

To change the default, constant output voltage greater than the input voltage can be obtained using boost converters or step-up converter. Boost converter employs a voltage feedback technique to step up the input voltage of variable or fluctuating nature to a higher constant DC output. Its main advantages are low losses high efficiency. The function of boost converter is enabled by the tendency of inductor to resist changes in current. The inductor plays the role of a load on the occasion of charging and during discharging, it takes the role of an energy source. The rate of change of current is the dependent factor by which the voltage produced during discharge phase is related (Fig. 9).

The input DC voltage ranging from 2 to 12 V, coming from the rectifier output, is to be stepped up to 14.5 V using the prescribed boost converter delivering up to 1.8 A.

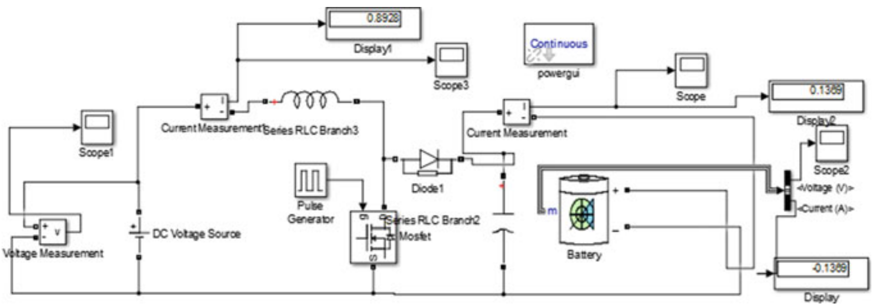


Fig. 9 Simulated circuit of boost converter in MATLAB

4.3.2 Buck Converter

The voltage supplied through the input is getting stepped down by using IC7805. This IC is suitable for low power applications. It steps downs from input of 12 V to output 5 V. Using this convertor with a heat sink to absorb the excess heat produced. The proposed system can be powered by a rechargeable lead-acid battery of 7 Ah and 12 V. Since the sum of currents of all the sensors, converters, microcontroller, GSM module and LCD displays is around 5 A, the prescribed capacity of the battery can power the system efficiently. This genre of battery can deliver high currents and has a remarkable tolerance to overcharging. Cost is low compared to other batteries having same rating.

5 Future Scope

Each and every user will have their own login ID and password where the account will be linked with a hospital; hence, they can cut the costs and develop a better relationship with the doctors. The system not only aimed to be implemented on bicycles but can be installed anywhere like wheelchairs, vehicles and even used as it is. Blood can be reached to a person in need on time as the system will be linked to all users being able to notify those with the same blood group. It becomes easy to locate people in critical condition; hence, reaching hospital in time for treatment cause system is linked with net and alert messaging also becomes useful.

6 Conclusion

This system poses to be eco-friendly, economic, less power and socially oriented mainly targeting the common people of all age groups. Aimed at keeping each one health-conscious aware and helping those in need to receive their health parameters before it is too late. The advantage of being able to instal this system not only in a cycle brings about taking into consideration the comfort of the user. The health of patients can be monitored day to day. Proposed method is an easy way of health monitoring in real time and communication between the doctor and the patient. Energy is fully utilized as to the cycle implementation being for patients who are obese to be able to lose weight and keep track of their health. All the above after being monitored physicians can access this through the web and give the required instructions then needed to be followed. Travel and time are reduced with this system.

References

1. Centers for Disease Control and Prevention (CDC) (2013) Vital signs: avoidable deaths from heart disease, stroke, and hypertensive disease-United States, 2001-2010. MMWR. Morbidity and mortality weekly report 62.35, pp 721
2. Aboughaly AA, Iqbal D, Abd El Ghany MA, Hofmann K (2017) NICBPM: non-invasive cuff-less blood pressure monitor. In: 2017 29th international conference on microelectronics (ICM)
3. Ray I, Alangot B, Nair S, Achuthan K (2017) Using attribute-based access control for remote healthcare monitoring. In: 2017 fourth international conference on software defined systems, pp 137–142
4. Wang L-H, Hsiao Y-M, Xie X-Q, Lee S-Y (2016) An outdoor intelligent healthcare monitoring device for the elderly. IEEE Trans Consum Electron 62(2):128–135
5. Anudev J, Jacob Raglend I (2012) Analytical study of GIC based current source model. In: IEEE international conference on advances in engineering, science and management, pp 219–222
6. Anudev J, Jacob Raglend I (2012) Analytical study of Howland current source model. In: 2012 international conference on computing, electronics and electrical technologies, pp 314–318
7. Karthik Reddy G, Lokesh Achari K (2015) A non invasive method for calculating calories burned during exercise using heartbeat. In: IEEE sponsored 9th international conference on intelligent systems and control (ISCO)
8. Czerwinski D, Wojcicki P, Zientarski T. On time blood pressure prediction with the use of PPG signals
9. Rajevenceltha J, Santhosh Kumar C, Anand Kumar A (2016) Improving the performance of multi-parameter patient monitors using feature mapping and decision fusion. In: 2016 IEEE region 10 conference (TENCON), pp 1515–1518
10. Sali S, Parvathi CS (2017) Integrated wireless instrument for heart rate and body temperature measurement. In: 2017 2nd international conference for convergence in technology (12CT)

A Fuzzy Rule-Based Diagnosis of Parkinson's Disease

D. Karunanithi and Paul Rodrigues

Abstract Neurodegenerative brain disorder is the root cause of Parkinson's Disease (PD). Neurodegenerative is the process of impairment of brain cells. PD is diagnosed through clinical methods. Hope this research work helps to identify the intensity of the PD. Fuzzy Inference System is used to identify the PD and its intensity. Fuzzy rules, Mamdani Fuzzy Inference, Membership Functions, and Defuzzification are the process used to obtain accurate results. Oxford Parkinson's Disease Detection Dataset is used for this research work. Among 23 fields in the dataset, only four fields FoH, DFA, Spread1, and Spread2 are chosen for analyzing the PD diagnosis and intensity. These four fields values are categorized into three sets: one is PD affected subjects, the second set is common values for both PD affected subjects and healthy subjects, and the third set is completely healthy subjects. Intensity values are measured from low to maximum as 0–100. Eighty-one rules are framed to calculate the PD intensity. We hope this FIS model is a novel method for identifying the PD intensity and helps doctors to diagnose and treat the patients in an effective way.

1 Introduction

Parkinson's Disease is a neurodegenerative brain disorder. When the brain reduces the secretion of the chemical called as neurotransmitter, neurotransmitter it directly impacts the body movements. This neurotransmitter is called as dopamine. The place where this chemical secretes in the brain is referred to as substantia nigra [1]. When these brain cells in the part of substantia nigra reduces or dies then it is said to be neurodegeneration [2].

D. Karunanithi (✉)
Computer Science and Engineering, Manonmaniam Sundaranar University,
Tirunelveli, Tamil Nadu, India
e-mail: karunanithid@gmail.com

P. Rodrigues
Computer Science and Engineering, King Khalid University, Abha, Saudi Arabia
e-mail: drpaulprof@gmail.com

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_116

In the eighteenth century, James Parkinson found the symptoms of this disease and wrote an article as “Shaking Palsy” [3]. PD affects the people after middle age and the ration is greater for male. PD is diagnosed by five stages from stage 1 to 5 as the severeness increases [4]. Electrocardiography [5] and voice signals are used for PD diagnosing [6–9]. Majority of the people with PD get affected with their voice [10–12]. So the speech recording dataset is taken for the processing and diagnosing the PD using Fuzzy Inference System.

2 Materials

Max Little from the University of Oxford created a dataset. This dataset consists of various voice recording measures of 31 people among 23 were PD affected and 8 were healthy. He created this dataset in collaboration with National Centre for Voice and Speech, Colorado. It consists of 23 columns of different voice measures. Each subject’s voice recordings are made five or six times and recorded the values in the respective column [13].

3 Methodology

3.1 Fuzzy Inference System (FIS)

Fuzzy Inference is a process of formulating the mapping from given input(s) to output(s) using fuzzy logic [14]. A fuzzy set is one in which the elements do not belong completely to only one set, but do belong to that sets to a certain extent [15]. An FIS was built on the three main components: fuzzifier, inference engine, and defuzzifier.

Figure 1 shows the detailed structure of the FIS. Normally, the database contains definitions such as information on fuzzy sets parameter with a function that has been defined for every existing linguistic variable. The basic rules can be constructed either from a human or automatic generation, where the searching rules using input–output data numerically.

There are several types of FIS, namely, Takagi–Sugeno, Mamdani, and Tsukamoto. An FIS of Mamdani is most commonly widely used methodology. It was proposed by Ebrahim Mamdani [Mam75].

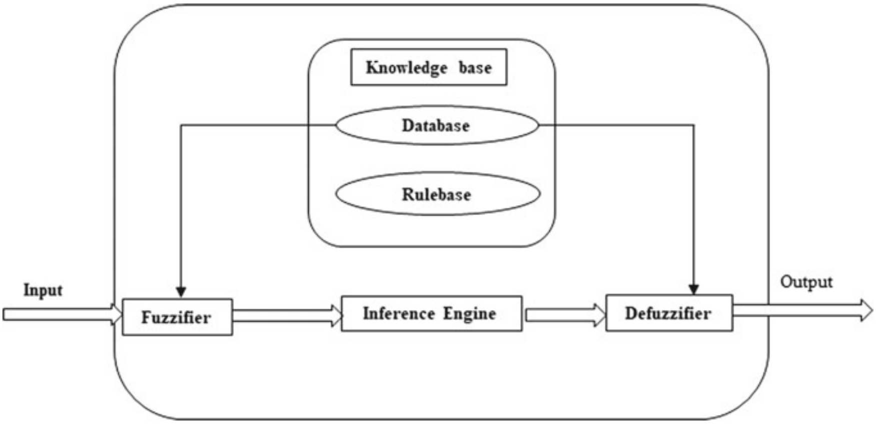


Fig. 1 Architecture of fuzzy inference system

3.2 Rules for Fuzzy Inference System to Diagnose PD

Sample fuzzy rules for predicting the status of PD are mentioned in Table 1. Data are taken from the Oxford Parkinson's Disease Dataset to frame these rules. Out of 23 fields in the PD dataset, only four fields are chosen for diagnosing PD using fuzzy logic. The four fields are FoH, DFA, Spread1, and Spread2. In the value analysis of the four fields, a clear difference between healthy and PD persons is observed. Also, these four fields correlate with my previous analysis of the same dataset using artificial intelligence. Three ranges had been identified in each field, which are a range of low values that completely fall with the PD affected persons, and mid range of values which belongs to both PD affected subjects and healthy subjects. And max. range of values completely belongs to healthy subjects. The middle range of values which belongs to both categories falls the fuzziness. Eighty-one rules are framed with these four fields to calculate the Parkinson's Disease Intensity (PDI) using fuzzy rules. PDI is calculated by nine ranges of values starting from 0, 12.5, 25, 37.5, 50, 67.5, 75, 87.5 to 100.

3.3 Fuzzification and Defuzzification

FIS designed for diagnosing Parkinson's Disease uses Mamdani Inference System. The defuzzifier uses the famous Center of Gravity (COG) method [16]. The formula for COG is

$$x* = \frac{\int \mu A(x).x\mathrm{d}x}{\int \mu A(x).\mathrm{d}x} \tag{1}$$

Table 1 Fuzzy rules for diagnosing PD

Rules for fuzzy inference system for diagnosis of Parkinson's disease
If (FoAvg is High) and (DFA is Low) and (Spread1 is Low) and (Spread2 is Low), then (PDI is Healthy)
If (FoAvg is High) and (DFA is Low) and (Spread1 is Low) and (Spread2 is Medium), then (PDI is VeryLow)
If (FoAvg is High) and (DFA is Low) and (Spread1 is Low) and (Spread2 is High), then (PDI is Low)
If (FoAvg is Medium) and (DFA is Medium) and (Spread1 is High) and (Spread2 is Low), then (PDI is Avg)
If (FoAvg is Medium) and (DFA is Medium) and (Spread1 is High) and (Spread2 is Medium) then (PDI is BAvg)
If (FoAvg is Medium) and (DFA is Medium) and (Spread1 is High) and (Spread2 is High), then (PDI is High)
If (FoAvg is Low) and (DFA is High) and (Spread1 is High) and (Spread2 is Low) then (PDI is High)
If (FoAvg is Low) and (DFA is High) and (Spread1 is High) and (Spread2 is Medium), then (PDI is VHigh)
If (FoAvg is Low) and (DFA is High) and (Spread1 is High) and (Spread2 is High), then (PDI is Max)

4 Experiment Results

4.1 Membership Functions to Diagnose PD

Mamdani FIS is used to diagnose the Parkinson's Disease. Four input member functions FrequencyAverage, DFA, Spread1, and Spread2 and one output member function Parkinson's Disease Intensity (PDI) are used in this experiment.

4.1.1 Membership Function for Frequency Average

From the analysis of the field Frequency Average (FoH), there are six recordings for each subject. The PD subjects and healthy are separated into two columns, and each subject's values are averaged and sorted in ascending order. After sorting, the three ranges of values are categorized as Frequency Average High (FoH), Frequency Average Medium (FoM), and Frequency Average Low (FoL). The range of values identified by the comparison of the healthy and PD subjects frequency average is FoH: 223.66 to 243.81, FoM: 114.30 to 203.90, and FoL: 97.94 to 113.01 (Fig. 2).

In this, FoH range of values belongs to healthy subjects. FoM range of values belongs both to healthy and PD affected subjects. Fuzziness observed in these ranges of values and another category FoL range of values belong completely to PD affected subjects. Three member functions were created to represent these values.

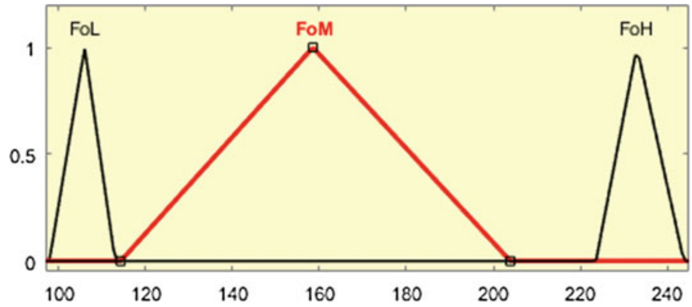


Fig. 2 Membership function for frequency average

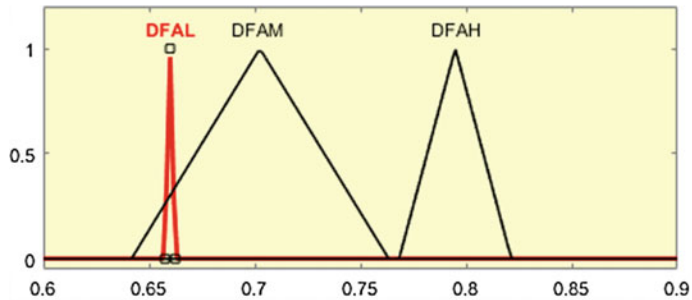


Fig. 3 Membership function for DFA

4.1.2 Membership Function for DFA

DFA field values are analyzed in the similar way FoH and categorized as DFA Low (DFAL), DFA Medium (DFAM), and DFA High (DFAH). The observed range of values for DFAL: 0.6383 to 0.6416 belongs to healthy subjects, DFAM: 0.6417 to 0.7634 belongs to both healthy and PD affected subjects where fuzziness occurs, and DFAH: 0.7686 to 0.8213 belongs to PD subjects. The membership functions for the above three are made and shown in Fig. 3.

4.1.3 Membership Function for Spread1

Spread1 field is analyzed in a similar pattern and categorized as Spread1 Low (SP1L), Spread1 Medium (SP1M), and Spread1 High (SP1H).

The range of values observed in SP1L: −7.59 to −6.70 belongs to healthy subjects, SP1M: −6.88 to −5.99 belongs to healthy and PD subjects, and SP1H: −5.63 to −3.66 belongs to PD subjects. The membership functions for Spread1 categories are made through FIS and shown in Fig. 4.

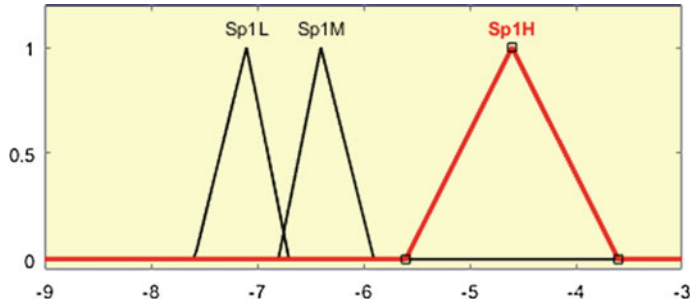


Fig. 4 Membership function for Spread1

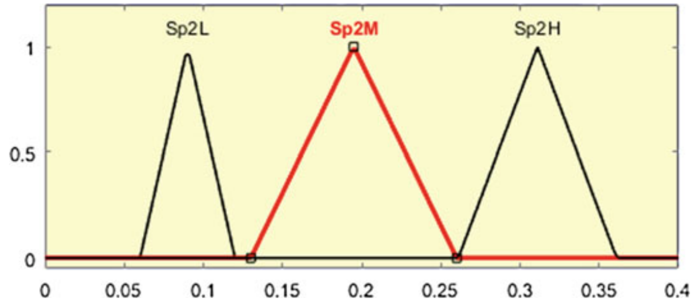


Fig. 5 Membership function for Spread2

4.1.4 Membership Function for Spread2

Spread2 field is analyzed in a similar pattern and categorized as Spread2 Low (SP2L), Spread2 Medium (SP2M), and Spread2 High (SP2H). The range of values observed in SP2L: 0.06 to 0.12 belongs to healthy subjects, SP2M: 0.13 to 0.26 belongs to healthy and PD subjects, and SP2H: 0.26 to 0.36 belongs to PD subjects. The membership functions for Spread2 categories are made through FIS and shown in Fig. 5.

4.1.5 Membership Function for Parkinson's Disease Intensity (PID)

From the four inputs FoH, DFA, Spread1, and Spread2, 81 rules are created to calculate the PDI with eight categories. The eight categories are mapped with the member functions Healthy (HTY), Parkinson's Disease Intensity Very Low (PDVIL), Parkinson's Disease Intensity Low (PDIL), Parkinson's Disease Intensity Below Average (PDIBAvg), Parkinson's Disease Intensity Average (PDIAvg), Parkinson's Disease Intensity Above Average (PDIAAvg), Parkinson's Disease Intensity High (PDIH), and Parkinson's Disease Intensity Very High (PDIVH).

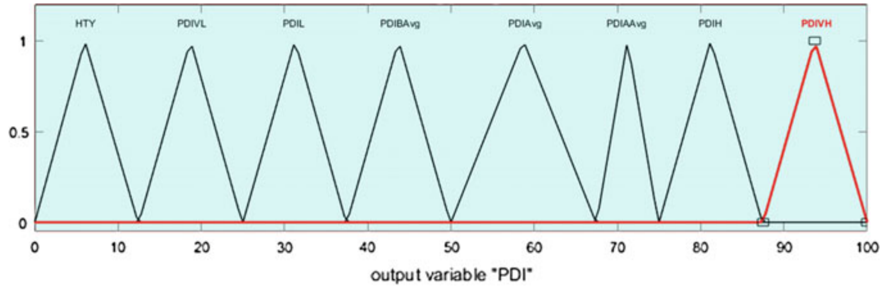


Fig. 6 Membership function for PDI

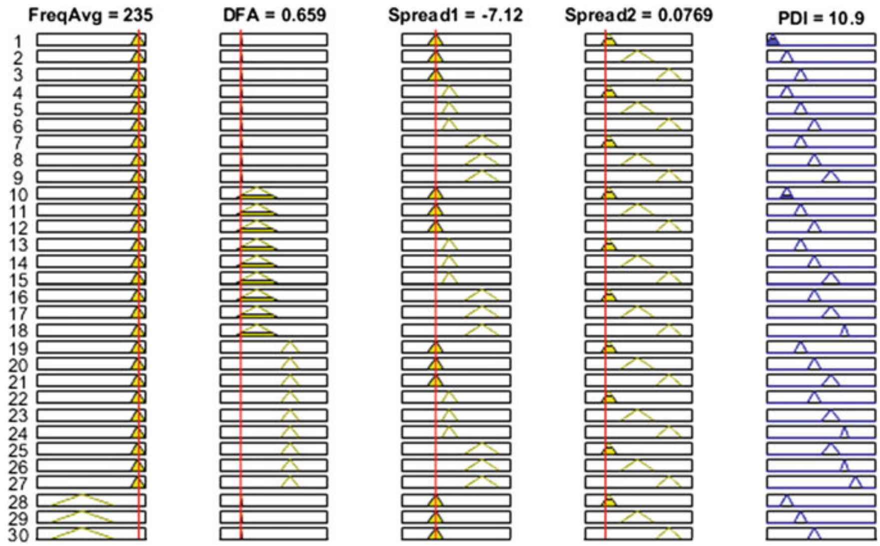


Fig. 7 Representation of fuzzy inference system for HTY

The measurements fixed for the eight member functions are HTY: 0 to 12.4, PDIVL: 12.5 to 24.9, PDIL: 25 to 37.4, PDIBAvg: 37.5 to 49.9, PDIAvg: 50 to 67.4, PDIAAvg: 67.5–74.9, PDIH: 75–87.4, and PDIVH: 87.5–100. Through the four inputs, various levels of intensity of the disease are calculated from the voice recordings of healthy and PD subjects. This fuzzy is a new method for identifying the intensity of the disease. The results will be more accurate and much useful for the doctors to treat the patients beyond the stages identification. The output member function is shown in Figs. 6, 7, 8 and 9.

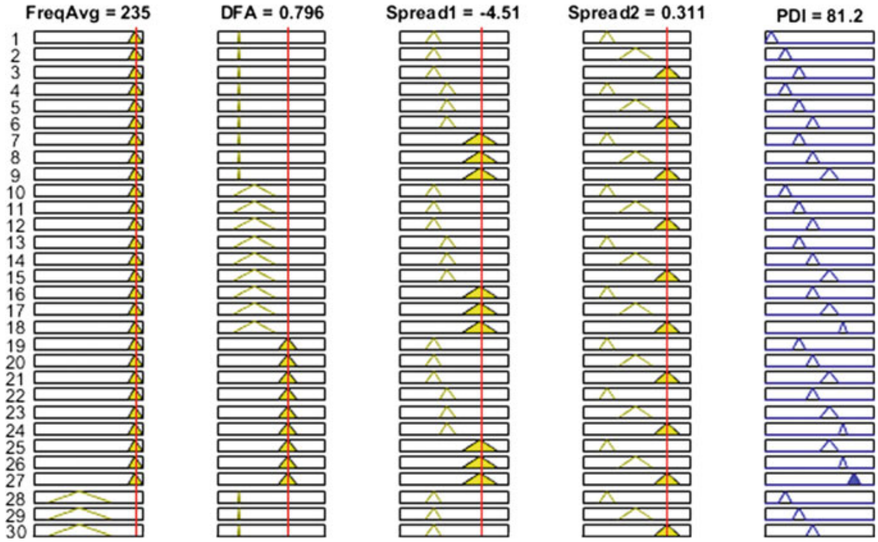


Fig. 8 Representation of fuzzy inference system for PDH

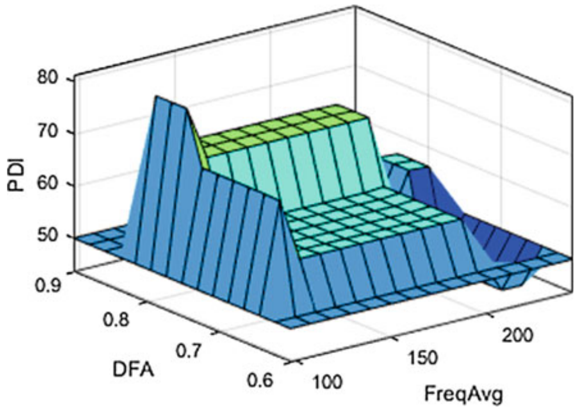


Fig. 9 Surface graph PDI with respect to DFA and FreqAvg

5 Conclusion

PD normally diagnosed by identifying the physical symptoms. Five stages of PD were identified by the severeness of the symptoms. In this research paper, the intensity of the PD was identified with eight levels from PD Oxford Dataset. In the value analysis of 23 fields, the chosen four fields, FoH, DFA, Spread1, and Spread2, values clearly show the difference between PD affected subject and healthy subjects. Also, few values are common for both subjects. Fuzziness occurs in these middle values of both subjects. Using fuzzy inference system, the intensity of the PD was able to be identified clearly. Eight levels of PD intensity were predictable through various possible combinations of four column values. With the four fields, tree structure was made with all possible combinations of four fields with the categories healthy, PD affected, and middle common values of between healthy and PD subjects. The intensity is calculated with the values 12.5 for fuzzy values of healthy and PD subjects and named as PPDV and 25 for complete PD affected values category and named as PDV. If any of the four fields values completely falls under the PD affected values, then the value of 25 is assigned. If any of the four field values falls under the middle category values that is the values which are common for both healthy and PD affected subjects 12.5 value is assigned. Eighty-one combinations were made with tree structure, and 81 rules are framed for each path of the tree. Using this PDV and PPDV, the final intensity was calculated from the range of 0–100. Eight categories of intensities: HTY, PDVIL, PDIL, PDIBAvg, PDIAvg, PDIAAvg, PDIH, and PDIVH were able to be obtained accurately through this FIS. This methodology might be useful for doctors to know the intensity of PD and treat the people accordingly.

References

1. Woo Y, Lee J, Hwang S, Hong CP (2013) Use of an adaptive-neuro fuzzy inference system to obtain the correspondance among balance, gait, and depression for Parkinson's disease. J Korean Phys Soc 62(6):959–965
2. http://www.parkinson.org/understanding-parkinsons/what-is-parkinsons. [updated in 2018]
3. Gelb D, Oliver E, Gilman S (1999) Diagnostic criteria for Parkinson disease. Arch Neurol 56(1):33–39. https://doi.org/10.1001/archneur.56.1.33.PMID9923759
4. http://www.parkinson.org/understanding-parkinsons/what-is-parkinsons/The-Stages-of-Parkinsons-Disease
5. Pezard L, Jech R, RuÊzÏicÏka E (2001) Investigation of non-linear properties of multichannel EEG in the early stages of Parkinson's disease. Clin Neurophysiol 122:38–45
6. Ene M (2008) Neural network-based approach to discriminate healthy people from those with Parkinson's disease. Math Comput Sci Ser 35:112–116
7. Caglar MF, Cetisli B, Toprak IB (2010) Automatic recognition of parkinson's disease from sustained phonation tests using ANN and adaptive neuro-fuzzy classifier. J Eng Sci Des 1:59–64
8. Gil D, Johnson M (2009) Diagnosing Parkinson by using artificial neural networks and support vector machines. Glob J Comput Sci Technol 9:63–71
9. Duffy RJ (2005) Motor speech disorders: substrates, differential diagnosis and management, 2nd edn. Elsevier Mosby, St. Louis
10. Ho AK, Iansek R, Marigliani C, Bradshaw JL, Gates S (1998) Speech impairment in a large sample of patients with Parkinson's disease. Behav Neurol 11:131–137. https://doi.org/10.1155/1999/327643
11. Logemann JA, Fisher HB, Boshses B, Blonsky ER (1978) Frequency and co-occurrence of vocal-tract dysfunctions in speech of a large sample of Parkinson patients. J Speech Hear Disord 43:47–57
12. Sapir S, Spielman JL, Ramig LO, Story BH, Fox C (2007) Effects of intensive voice treatment (the Lee Silverman Voice Treatment [LSVT]) on vowel articulation in dysarthric individuals with idiopathic Parkinson disease: acoustic and perceptual findings. J Speech Lang Hear Res 50:899–912. https://doi.org/10.1044/1092-4388(2007/064)

13. Center for Machine Learning and Intelligent Systems (2008). http://archive.ics.uci.edu/ml/datasets/Parkinsons
14. Khezri R, Hosseini R, Mazinani M (2014) A fuzzy rule-based expert system for the prognosis of the risk of development of the breast cancer. Int J Eng (IJE) Trans Basics 27(10):1557–1564
15. Camara C, Warwick K, Bruña R, Aziz T, del Pozo F, Maestú F (2015) A fuzzy inference system for closed-loop deep brain stimulation in Parkinson's disease. J Med Syst 39(11)
16. Hamidzadeh J, Javadzadeh R, Najafzadeh A (2015) Fuzzy rule based diagnostic system for detecting the lung cancer disease. J Renew Nat Resour Bhutan 3(1):147–157

A Comprehensive Study of Retinal Vessel Classification Methods in Fundus Images for Detection of Hypertensive Retinopathy and Cardiovascular Diseases

J. Anitha Gnanaselvi and G. Maria Kalavathy

Abstract Quantitative studies for classification of retinal vessels using new computer-assisted retinal fundus imaging system have allowed the researchers to understand the influence of systemic on retinal vascular caliber. These retinal vascular caliber changes reflect the cumulative response to cardiovascular risk factor. Hypertensive retinopathy can be detected in earlier stage by analyzing the retinal image. Nowadays, it is obvious that there is a relationship between changes in the retinal vessel structure and the most common diseases such as hypertension, stroke, cardiovascular diseases, those can be detected by noninvasive retinal fundus image. The proposed approach of applying an image processing technique, the aforementioned disease can be diagnosed earlier by retinal fundus image. To achieve the precise measurement of the retinal image parameters, the classification of blood vessels such as arteries and veins is necessary. These classifications of arteries and veins can be achieved through the retinal fundus image. The retinal vessel classification is based on visual and geometric features from these classified images into arteries and veins for the detection of hypertensive retinopathy, stroke, and cardiovascular risk factor. This classification of retinal fundus image is essential for early diagnosis of aforementioned diseases. The retinal arteriolar caliber which is narrower and smaller, that is associated with older age, will predict the incidence of diabetic retinopathy and cardiovascular risk factor. Similarly, retinal venular caliber which is wider, that is associated with younger age, will predict the incidence of risks of stroke and coronary heart diseases. This could suggest the possibility of using this model of fundus image in classification approaches. Finally, the selected attributes of classification are applied through the genetic algorithm with radial basis function neural network for diagnosis of the disease in order to improve the classification accuracy with less computational cost time.

J. A. Gnanaselvi (✉)
Research Scholar, Faculty of Information and Communication Engineering, Anna University, Chennai, India
e-mail: anithagnanaselvi@gmail.com

G. M. Kalavathy
Department of CSE, St.Joseph's College of Engineering, Chennai, India
e-mail: maria_kalavathy@yahoo.co.in

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_117

1 Introduction

Diabetic retinopathy (DR) means common problem of diabetes which damages the retinal vascular area. Mostly, it affects the blood vessels in retina. One of the major issues in diabetic retinopathy is visual impairment. This is because of the new blood vessel growth in retina in proliferative retinopathy. There are several categories to predict the early diagnosis of diseases. One of the best methods is to predict the DR by using the fundus image, this fundus image is considered to be the initial and basic screening process in diabetic retinopathy prediction.

Among the several blood vessels, the retinal blood vessel network is visible and is suitable for noninvasive imaging method in our body. So, the retinal blood vessel is considered to be the reliable tool for early diagnosis of DR. In order to perform the analysis, the classification of blood vessel is done based on geometric features of vessel network. In retinal image analysis, the accurate measurement of retinal vessel parameters is considered to be an important problem in eye research. There are several parameters measured from the retinal vessel. This includes the thickness and curvature of the vessels and also the measurement of the arteriolar–venular ratio.

The arteriolar–venular ratio is considered as the essential parameter for early prediction and diagnosis of diseases which includes the hypertension, stroke, cardiovascular disease in youngsters, and retinopathy in child. Though there are several sets of rules that have been defined for measuring the ratio of arteriolar–venular. These include the distance from the optic disk margin; in Japan, the ratio is measured and calculated as 0.25–1 of the optic disk diameter. In the U.S., it is stated as 0.5–1 for measuring the arteriolar–venular ratio calculation.

This arteriolar–venular ratio calculation measurement includes other areas also such as localization of optic disk, retinal vessel diameter, measuring accurately, retinal vessel network image analyzing, and also for the classification of blood vessels which includes arteries and veins. This classification of blood vessel is considered as the basic step for calculating the arteriolar–venular ratio. This arteriolar–venular ratio is important for calculating the classification of blood vessels in an efficient and effective way.

The quantitative calculation of retinal vascular caliber is highly influenced in the clinical significant association outcomes such as detection of stroke and coronary heart diseases. This can be achieved by the retinal fundus image.

2 Related Work

In the modern world, by the early diagnosis many diseases can be controlled that includes the life-threatening diseases. Though there are several approaches that are existing already for the early diagnosing diseases, this deals with the two categories. First category used the machine learning popular algorithm. Genetic algorithm which is used for the feature selection of the attributes is included in the second category and it uses the radial basis functional neural network for the classification purpose in the attributes. The diabetic retinopathy diagnosed by the Pima Indian Diabetes Dataset is used for the classification purpose. The results show that it will minimize the cost of computation time and is also better in classification [1, 2].

The diabetic retinopathy (DR) mainly affects the retinal vessels in the macular region which is located in the center portion of the retina in the fundus image. These studies are based on the four main methods. It contains the preprocessing of the image and then enhancement of the fundus image. Third is the segmentation process which will segment the vessels, and finally it includes the proposed foveal avascular zone segmentation approach which is used to diagnose the diseases. This shows the results of an average in performance metrics such as specificity, sensitivity and accuracy from theses diseases are diagonsed earlier [3, 4].

There are various approaches to quantify the width and tortuosity of retinal vessels to detect cardiovascular diseases. The artery-to-venous ratio will predict the narrow arterial and venous dilatation for the detection of stroke and cardiovascular heart diseases. Here, the artery-to-venous classification method is used which provides the accurate region around the optic disk [5].

The segmentation of retinal vascular blood vessels into veins and arteries is used for quantifying the ratio of artery-to-venous diameter. It is developed to predict the cardiovascular heart disease and stroke in children. The features are extracted from the databases and used for the training and testing phase algorithm for all the vessels and it will point out whether it is vein or artery. The artery-to-vein classification approach using the receiver operator characteristic shows better results and has detected the artery and vein in the fundus image [6].

In order to achieve the accurate calculation and measurement of parameter in diagnostic features, the arteries and veins are essential. The classification parts of blood vessels in the fundus image are evaluated by different databases in different criteria. These approaches focus on the geometric features and statistical model in spatial and transform domain [7, 8].

3 Methodology and Measurement

The common method of the blood vessel classification is by utilizing the fundus image. First step is segmentation of vessels, second step is selecting the region of interest for classifications of vessels. Third step is extracting the features from various parts of the vessels. Fourth step is to classify the vectors of features. At last, final is the combination of results to determine the final label of vessel. This is considered as the existing way to classify the vessels in the fundus image.

The proposal is based on the classification of vessels into two major criteria. They are represented as automatic and semiautomatic methods.

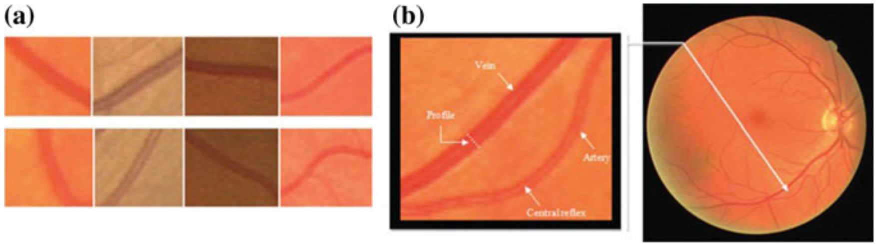


Fig. 1 **a** Sample images of vein represented in first row and artery represented in second row in fundus images, **b** Specifying central reflex and profile in a piece of a fundus image

3.1 Semiautomatic Method

The semiautomatic method for retinal vessel analyzes the artery and venous individually and it calculates geometrical and topological pixel features of each segment. This method uses the anatomical properties of veins and arteries. The semiautomatic method is a hybrid of graph-theoretic method with domain-specific knowledge and is capable of analyzing the entire vasculature (Fig. 1).

3.2 Automatic Methods

In this automatic approach, position around the optic disk is within 0.5–2 diameter of disk from its center portion and is segmented into four zones, in which each one contains one of the major arches. In these, the red channel and hue channel in each vessel segment are represented as the most discriminative features for classification. As the results, in two adjacent vessels, the darker reddish vessel is considered as vein, the vessel that has more color uniformity is considered as vein (Fig. 2).

After feature extraction, the classification criterion of vessels is based on fuzzy clustering algorithm. The Euclidean distance of each pixel from the mean value of features is calculated. After vessel classification, the data is available by vessel identifying the variation in pixels to discriminate the arteries and veins. Twenty-six fundus images have been taken in this study for assessment, in which ten fundus images were used to develop the algorithm, and 16 images were used for validation. Reported results on 15 validation images show the overall error of 12.4%. This classification procedure is performed around the optic disk and it is more preferable for optic-disk-centered images.

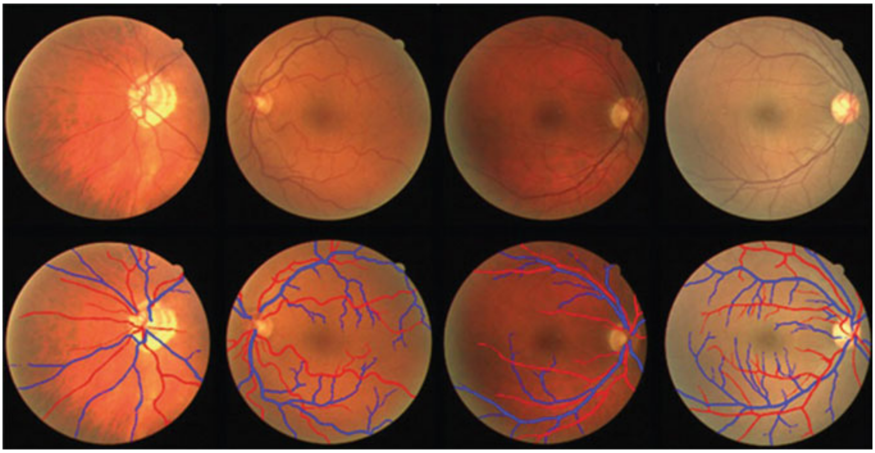


Fig. 2 First row represented some samples of fundus images from dataset and second row represented their ground truth. The arteries are represented in red lines, and the veins are represented in blue lines

3.3 Retinal Vessel Measurements

The retinal vessel calibration is based on the computer-assisted program to find out the accurate size of the fundus image. The calculation consists of actual size of the central retinal artery, the same procedure is applicable for vein, and it calculates the ratio of these abovementioned variables which is known as the artery-to-vein ratio (AVR). The result is obtained and is represented as shown in Fig. 3.

The measurement of retinal vascular caliber formula is implemented based on the theoretical and empirical methods and it can be achieved by larger quantity of retinal image with its branch point. The measurement of individual trunk vessel and branch vessel is obtained by square root mean deviation model for observing the data.

In today's population, it is difficult to find out the problem such as hypertension diabetes and ocular alignments problem such as diabetic retinopathy and glaucoma [9]. Those diseases can be measured by the retinal vessel through fundus Image. This can be measured through the refraction/axial length and retinal vascular caliber by retinal AVR (artery-to-venous) ratio calculation. However, it is very essential to identify the narrower artery and wider venous and this will result in the smaller artery-to-venous ratio. Figure 3 represents the arteriolar–venular ratio calculation measurement zones.

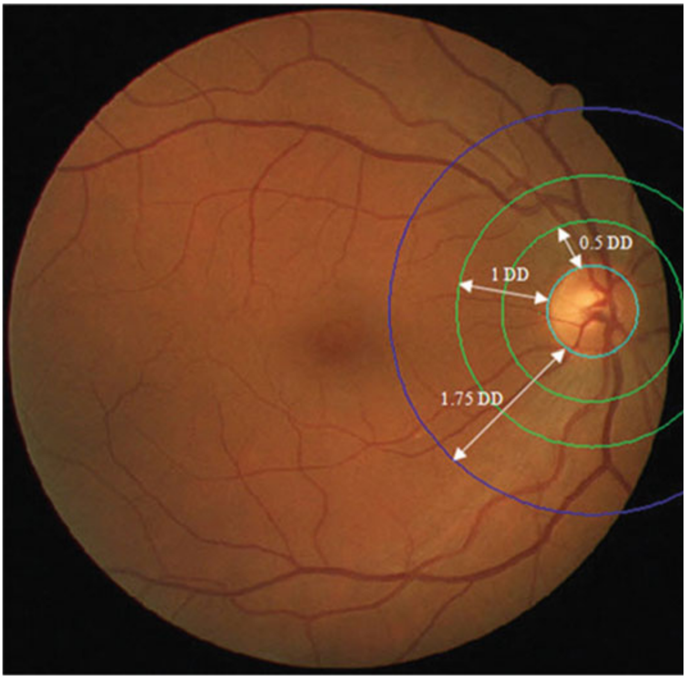


Fig. 3 Arteriolar–venular ratio (AVR) calculation measurement zones, DD = optic disk diameter

4 Results and Discussion

4.1 Proposed Methodology Using GA with RBF NN Classification of Diseases

The proposed methodology is implemented with genetic algorithm, with radial basis neural network function for classification of attributes. Here, we used Pima Indian Diabetes Dataset to perform the best diagnosis of diabetes patients in datasets. Here, the proposed method in this abovementioned algorithm is implemented to detect the diabetic patients in Pima Indian Diabetes Datasets. The genetic algorithm based on selection, crossover, and mutation operation is performed in order to obtain better attribute features and it achieves the shorter in training and testing time, that is, computational cost and time, better storage capacity, and finally increases the classification rate. Table 1 represents the proposed algorithm architecture.

The radial basis functional neural network is based on supervised learning approach. In this proposed method, the radial basis functional neural network has been used for the classification in order to diagnose the diabetes disease (Fig. 4).

The radial basis functional neural network is a supervised feedforward process which consists of one hidden layer of hidden units, which states the radial basis

Table 1 Proposed algorithm

Step1: Initialization
Step2: Store the Pima Indian Diabetes Dataset
Step3: Start the parameters of genetic algorithm
Step4: Run the genetic algorithm
Step5: (a) First generation process starts
Step6: Selection While stopping criteria not met do
(b) Crossover
(c) Mutation
(d) Selection
End
Step6: Applying the radial basis functional neural network for classification
Step7: Dataset attributes are trained
Step8: Calculation of accuracy and measuring the error
Step9: Testing the datasets
Step10: Calculation of accuracy and measuring the error
Step11: Stop the process

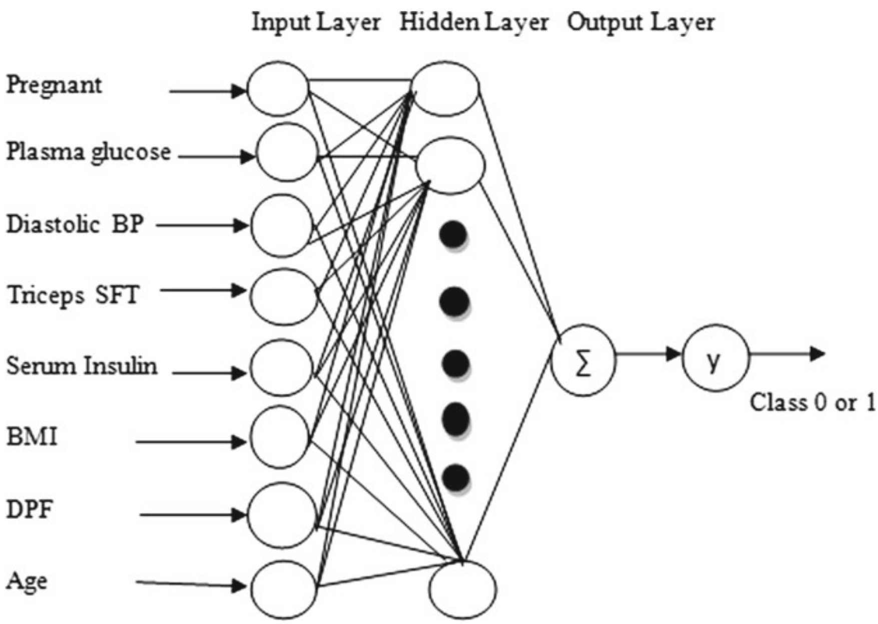


Fig. 4 Feedforward neural network model for diagnosis of diabetes

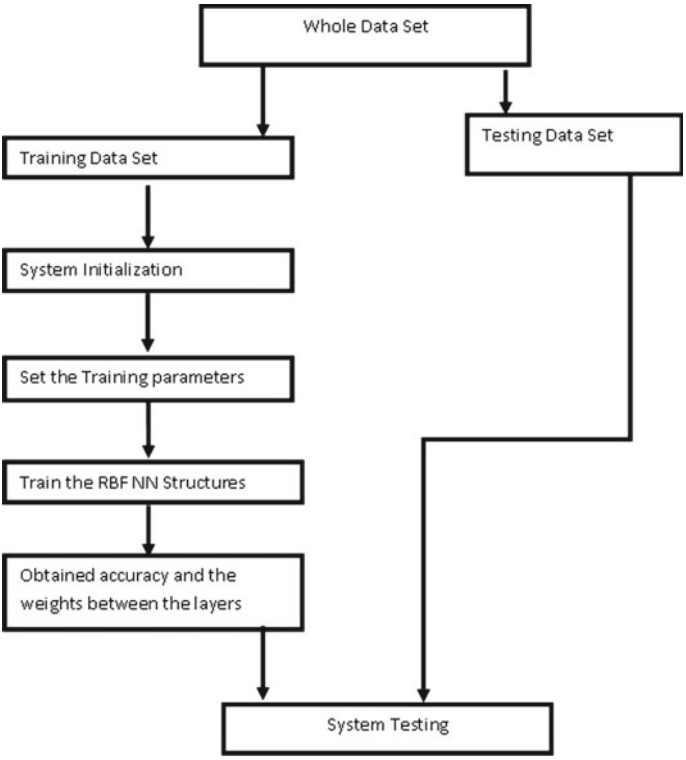


Fig. 5 The radial basis functional neural network methodology

Table 2 Genetic algorithm attributes selection

Dataset	Number of attributes	Name of the attributes	Number of instances	Number of classes
Pima Indian Diabetic Datasets with genetic algorithm	1	Plasma glucose tolerance test	768	2
	2	Serum insulin		
	3	Body mass index		
	4	Age		

functions (RBFs). These radial basis functional neural networks require a desired response to be trained in pattern classification studies. Particularly, in the present study, a training algorithm is used which normally uses a gradient descend rule for the training attributes of these networks. Figure 5 represents the radial basis functional neural network methodology (Figs. 6, 7 and Tables 2, 3).

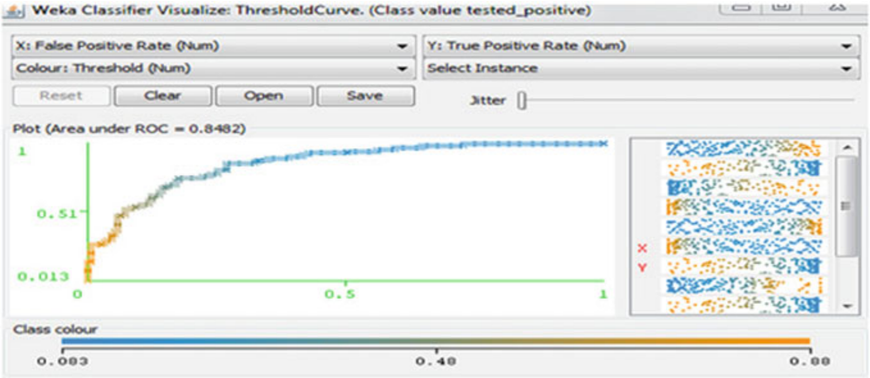


Fig. 6 ROC graph GA_RBF NN on PIDD

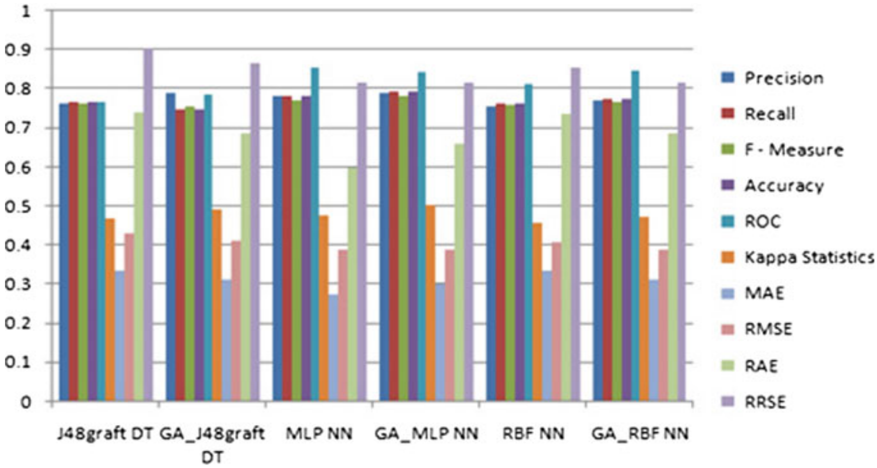


Fig. 7 Evaluation GA_RBF NN performance for PIDD

5 Conclusions and Future Directions

Diabetic blood sugar when above the desired level is considered as the world's widespread disease. The contribution of this proposed method was developed and implemented with genetic-algorithm-based radial basis functional neural network. It also determines and assesses the diabetic classification and detection of artery and vein in order to estimate the various problems including the blindness, blood pressure, kidney diseases, coronary heart diseases, nerve damages, and so on. In this proposed study, firstly, the classification has been done on Pima Indian Diabetic Datasets using radial basis functional neural network, and then using genetic algorithm for attributes selection, and thereby performing classification on the selected attribute. The results obtained are more interesting and may also happen for the exploration of the dataset.

Table 3 Results of genetic algorithm radial basis functional neural network for Pima Indian Diabetic Datasets

Measure	Training set evaluation	Testing set evaluation
Precision	0.75	0.75
Recall	0.76	0.76
F-measure	0.756	0.77
Accuracy	0.766	0.76
ROC	0.819	0.80
Mean absolute error	0.55	0.49
Root mean squared error	0.41	0.42
Relative absolute error	0.71	0.72
Root relative squared error	0.85	0.86
Kappa statistics	0.45	0.46

These classifications of arteries and veins can be achieved through the retinal fundus image. The retinal vessel classification is based on visual and geometric features from these classified images into arteries and veins for the detection of hypertension, stroke, and cardiovascular risk factor. From this, we tend to conclude that the proposed approach will help physicians to improve or take accurate decisions to do work speedily with less expense. In future, these proposed approaches can also be used for other kinds of diseases.

Acknowledgements All the images are taken from "Pima Indian Diabetes Dataset" which is publicly available.

References

1. Choubey DK, Paul S (2017) GA_RBF NN: a classification system for diabetes. Int J Biomed Eng and Technol 23(1):71–91
2. Yadav P, Ruhil N (2016) Blood vessel detection for diabetic retinopathy. In: IEEE international conference. 978-9-3805-4421-2/16/$31.00
3. Nugroho HA, Dharmawan DA, Litasri (2017) Automated segmentation of foveal avascular zone in colour retinal fundus images. Int J Biomed Eng and Technol 23(1):1–18
4. Sun C, Wang JJ, Mackey DA, Wong TY (2009) Retinal vascular caliber: systemic, environmental and genetic associations. Surv Ophthalmol 54(1):74–94
5. Joshi VS, Reinhardt JM, Garvin MK, Adramoff MD (2014) Automated method for identification and artery-venous classification of vessel trees in retinal vessel networks. PloS 9(2):e88061. www.plosone.org
6. Niemeijer M, van Ginneken B, Abramoff MD (2016) Automatic classification of retinal vessels into arteries and veins. In: Proceedings of SPIE-the international society for optical engineering, 7260, 72601F. https://doi.org/10.11117/12.813826
7. Miri M, Amini Z, Rabbani H, Kafieh R (2017) A comprehensive study of retinal vessel classification methods in fundus images. J Med Signals and Sens 7(2):59–70

8. Guzman JC, Melin P, Prado-Arechiga G (2017) Neuro-fuzzy hybrid model for the diagnosis of blood pressure. In Nature-inspired design of hybrid intelligent systems. Springer International Publishing, pp 573–584. https://doi.org/10.1007/978-3-319-47054-2_37
9. Agarwal A, Williams GH, Fisher NDL (2005) Genetics of human hypertension. Trends in Endocrinol and Metab Elsevier. 10.10.16/j term

Estimation of Parameters to Model a Fabric in a Way to Identify Defects

V. Subhashree and S. Padmavathi

Abstract Fabric defect detection is a quality check process which can locate and identify defects caused during the production process in the textile industry. Automated defect identification system uses computer vision and pattern recognition techniques whose performance depends majorly on the quality and quantity of the input dataset. A wide range of parameters is considered for decision process which compromises the accuracy of the system. This paper aims to estimate suitable parameters for the defect-free fabric which can be used by traditional methods to identify the defects in an efficient way. Hough-transform-based method is proposed to identify the parameters and the algorithm is experimented on various fabrics. The proposed method gives promising results when the horizontal and vertical threads are evident in the image.

1 Introduction

The defect detection at an earlier stage is an essential requirement for the fabric production industries. The quality control of the fabric affects the duration of production and is getting high productivity without the financial loss. According to paper [1], the cloth rates are scaled down by 45–65% because of the fabric defects caused during production process. To improve the quality, preparatory measures can be taken in prior before the item reaches the market [2]. The defects are caused by error patterns of the yarns or the spoils made by the machines during the production. The discontinuities of the threads on the fabric are an essential thing to be identified during manufacturing process. The identification of the minute thread defects on the fabric using man power during the production is tiresome and delayed. The manual

V. Subhashree · S. Padmavathi (✉)
Department of Computer Science and Engineering, Amrita School of Engineering, Coimbatore, Amrita Vishwa Vidyapeetham, Coimbatore, India
e-mail: s_padmavathi@cb.amrita.edu

V. Subhashree
e-mail: cb.en.p2cvi16008@cb.students.amrita.edu

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_118

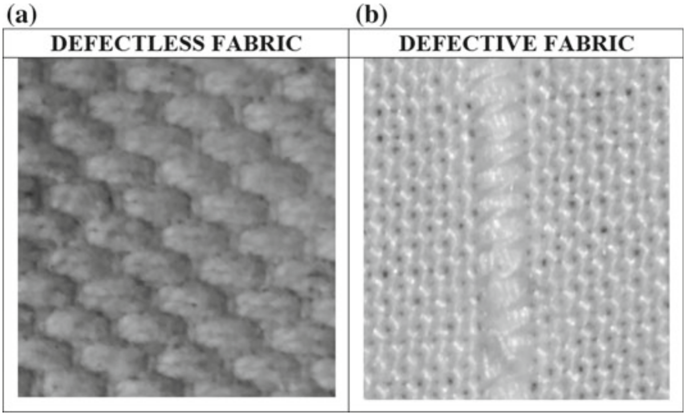


Fig. 1 **a** Defectless fabric, **b** defective fabric

method of defect identification is claimed to give recognizable results of only 70% with a limited speed. Hence, usage of automated defect identification systems helps manufacturing industries to complete the requirements within a specific time without compromising the quality.

Automated fabric defect detection is the need of the textile industries and image processing [3, 4] can be used as a solution for such problem. The automated inspection uses machine learning algorithms to identify the defects from the random patterns on the fabrics.

The algorithms use thread disorder identification in extensive range for defect detection on the fabric. The method indirectly utilizes the yarn dimensions like thickness, density, and other characteristics of the yarn. The algorithm should be able to accurately classify various types of defects on the fabric. The challenges of image-processing-based techniques include illumination difference, change in the orientation, occlusion of the yarns, etc.

It is quite common that the defects occur on different fabrics and is very less likely that all defects occur on a particular fabric. The change in the fabric becomes a factor that influences the defect detection system. When there is an increase in the input dataset, the variations on the fabric are also likely to increase. The accuracy of the defect detection system does not purely depend on the defect and in fact is much dominated by the fabric material. This phenomenon is demonstrated in Fig. 1 which shows two different fabrics with varying thread parameters. The defects occur at the thread level and the influence of these defects on the overall fabric image is very less which in turn lead to inaccuracy in classification.

There is an abundant availability of the defect-free fabric images than the defective images, which influences the dataset used for training the classifier. The above discussion demands a requirement of suitable parameters from the defect-free images which could be utilized for identifying the defects. The parameters like (i) distance between the threads, (ii) orientation of the thread, and (iii) density/thickness of the thread are few of the parameters which differ between the defective and defect-free fabric. In this paper, we attempt to model defectless fabric using the thread parameters measured from fabric images.

This paper is organized as follows: The literature survey is discussed in Sect. 2, proposed work is in Sect. 3, experimental analysis is in Sect. 4, and conclusion is in Sect. 5.

2 Literature Survey

The texture of the fabric is the most important strategy to model the defectless fabric. Texture methods used for representing defectless fabric can be majorly classified into (i) model based, (ii) spectral based, (iii) statistical, (iv) structural, and (v) hybrid and motif based [5]. GLCM and Gabor filter are most widely used applications for fabric defect detection. The relationship between the pixels is analyzed using the statistical approaches. The statistical analysis between the pixels is classified based on the first-order, second-order, and higher order methods. The gray-level co-occurrence matrix (GLCM) is categorized under the second order, which is the most predominantly used vision-based methods for defect identification on the fabrics [5]. Different features can be extracted from gray-level co-occurrence matrix as proposed by Harlick et al. that can be used to represent the texture properties as in [6] and patterns. Gabor filter is a widely used spectral method to represent the texture.

2.1 GLCM

The GLCM matrix has occurred by extracting the features which are based on second-order statistics. The features extracted by the first-order statistics are the gray-level scale distribution in the image, which does not contain the information about the different gray levels in the image. Jagdish Lal Raheja [7] proposed an effective method to extract the texture features using the distance and orientation parameters using GLCM method. The co-occurrence matrix is arranged with the corresponding frequencies of different gray levels $P\theta, d(I1, I2)$ to produce texture features. It measures the corresponding probability of the measured distances and the directions in a matrix [2].The co-occurrence matrix consists of two parameters: (1) d—distance measurement between the pixels and (2) θ—orientation between the pixels.

2.2 Gabor Filter

Gabor filter is the most useful application for identifying the edge detections, documents management, and image segmentation [7].

$$g(x, y; \lambda, \theta, \varphi, \sigma, \gamma) = \exp - ((x^{\wedge\prime}2 + \gamma^2 y^{\wedge\prime}2)/(2\sigma^{\wedge}2))\cos(2\pi\, x^{\wedge\prime}/\lambda + \varphi) \quad (1)$$

Gaussian function is modulated by Gaussian kernel function. Daugman [8] proposed a method by aggregating the properties on the receptive field areas, which is the visual cortex and the sinusoidal Gabor function is shown in Eq. 1 which is for definite frequency and orientation change. Padmavathi [9] had an approach to wavelet transform using Gabor filterbank from a large set of parameters. The Gabor consists of large sets of parameters [10] like aspect ratio, bandwidth, phase offset, orientation, and wavelength. The choice of optimal parameters is always a requirement. Evolutionary algorithms [11] are recently used for optimization problems. Tsaia [12] attempts to choose optimal Gabor parameters for texture segmentation. Daugman [13] tries to capture the spatial frequencies through Visual Cortical filters.

The morphological techniques are widely used for fabric inspections [11]. The morphological operations perform better on fabric defects with hole and missing of the yarns [14]. Priya [2] was able to achieve better results by performing morphological operations on the low-contrast images. Meihong Shi [4] reduced texture erroneous and noises on the image by performing the LCD method.

The Hough transform is useful for measuring the global descriptors of the features. Mid point Hough transform is suggested by Hari in [15] to increase the speed of line detection. The fabric consists of warp and weft threads on the fabric, by finding the skew direction of warp and weft it is easy to calculate the fabric density [3]. Yildiz [1] proposed a system of the thread densities detection using the horizontal and vertical frequencies of the fabric by computing FFT. Estimation of the fabric parameters is done using the Hough transform. Ruru Pan [16] used Hough to segment out the warp and weft in both directions to find the densities between the yarns.

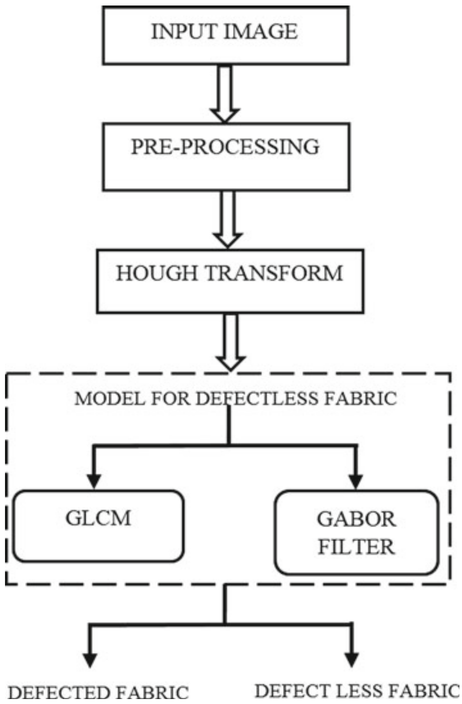
This paper aims to develop a model based on the thread parameters from the defectless fabric. The fabric defects depend on the thread parameters, the threads are restricted to warp and weft direction of the fabric. Hence, this paper concentrates on the parameters in these directions.

3 Proposed Work

Fabric defect detection process involving GLCM and Gabor filter usually computes the response for a large set of parameters and their average is used for defect detection. The small variations of the thread occurring in the defective fabric lose its importance when computed across wide range of parameters. When the parameters are narrow and closer to the thread parameters, the fabric defect detection accuracy can be improved. This paper estimates these parameters based on the thread parameters of a defectless fabric. The parameters estimated by the proposed method could be used to fine-tune the GLCM and Gabor filter parameters. The overall process of the proposed system is given in Fig. 2.

The input defectless image is converted into a binary image using threshold method. The binarized image is preprocessed using the morphological operations

Fig. 2 System architecture diagram



of erosion followed by double dilation to enhance the connectivity of thread parameter. The Hough transform is applied to the resulting image to detect the lines. Hough transform is used for line detection using the local measurement. The polar coordinate system is expressed as follows:

$$x \cos + y \sin = d \tag{2}$$

where r represents the distance between the origin and the line "θ" represents the direction with reference to "r". The (d, θ) values correspond to (xi, yi), which maps to Cartesian space in the Hough space. Hough transform accumulates the count of line parameters present in the image. Dominating peaks of Hough lines are considered as thread lines. These lines are converted to spatial coordinates and the distance between them is computed. These distances are considered as thread parameters of the fabric. Finally, the refined parameters are sent to model the defectless fabric. These parameters are used to fine-tune the GLCM parameters. The features like energy, homogeneity, entropy, contrast, and correlation are calculated. The procedure is experimented on different fabrics and the results are summarized.

Table 1 GLCM features

Features	Values d = 2, Theta = 30	Values d = 2, Theta = 90	Values d = 5, Theta = 90
Energy	0.4535	0.6524	0.6333
Homogeneity	0.9335	0.7542	0.9666
Correlation	0.7826	0.8562	0.8125
Contrast	0.1332	0.1265	0.669

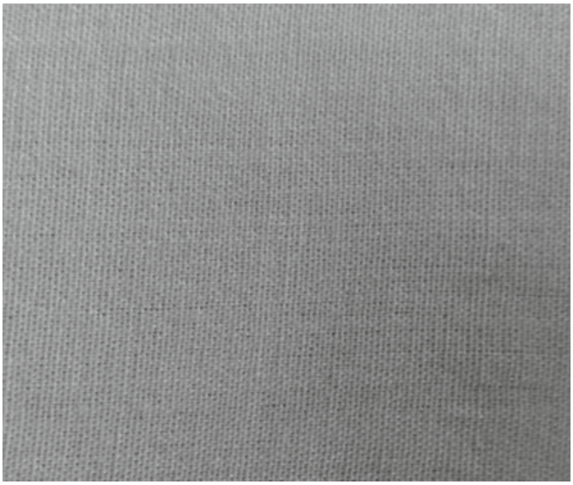


Fig. 3 Input image

4 Experimental Analysis

The dataset consists of 35 set of images, which includes the images at three different orientations 0,120, and 45° of camera. The dataset consists of 14 defectless fabrics, 21 defective fabrics, and four different thread thickness/thread patterns. Table 1 shows the GLCM features of the same image calculated for three different parameter sets. It could be observed that the values vary across the parameters, thus justifying the need for fine-tuning the parameters.

The input image shown in Fig. 3 is preprocessed as explained in previous section, and the resulting image is shown in Fig. 4. It is clear from the image that the morphological operations increase the visibility of the warp and weft yarn in the fabric.

Hough transform is applied to the preprocessed image and the peaks are calculated as explained in the previous section. Those dominating peaks are converted to lines and shown on the image as in Fig. 5. The distance between the lines was calculated using the dominating peak coordinates.

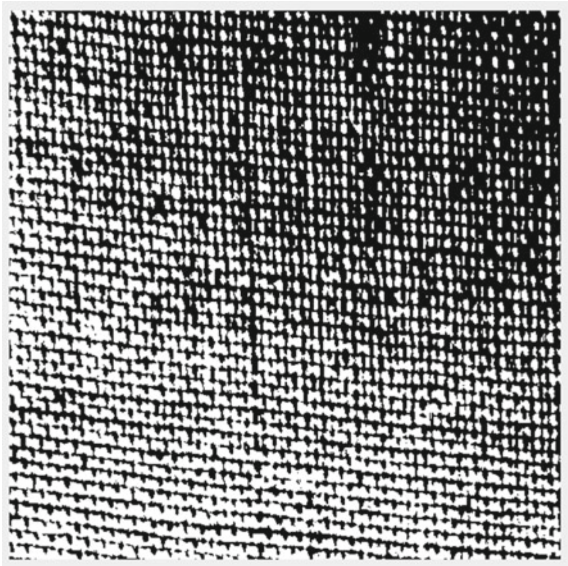


Fig. 4 Preprocessing

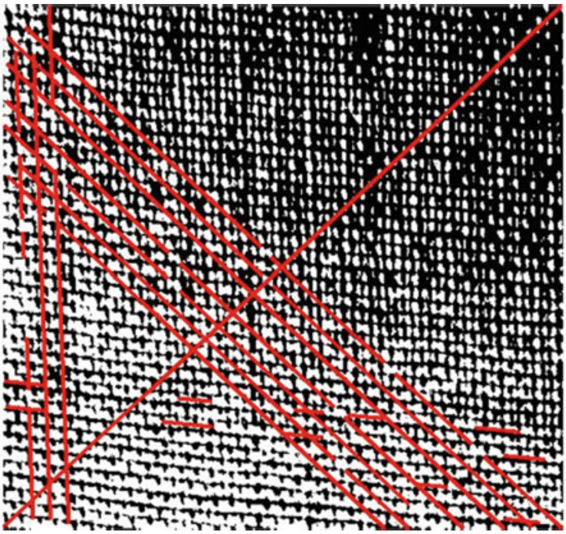


Fig. 5 Hough transform

The distance between horizontal lines is computed as 51 and distance between vertical line is computed as 38 for the test image shown. From the observed distances, the values are used for calculating the GLCM features.

From the graph, it could be observed that the calculated value is close to the observed peak. The texture correlation for the horizontal offset is shown in Fig. 6. Since the vertical thread is more enhanced than the horizontal threads in the preprocessing stage, the proposed method highlights approximately two-thread distance

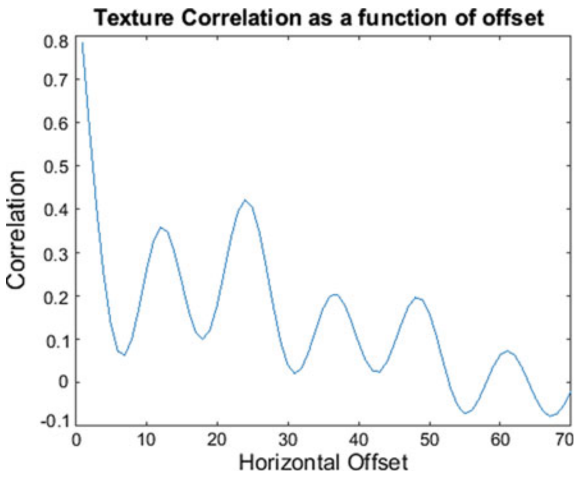


Fig. 6 Horizontal offset

Table 2 GLCM feature calculation for 10 images

Image	GLCM Features				
	Homogeneity	Energy	Correlation	Entropy	Contrast
1	0.8924	0.5964	0.8841	0.8641	1.4512
2	0.8645	0.6095	0.8990	0.9745	1.3774
3	0.9245	0.5421	0.9321	0.7541	1.0121
4	0.9214	0.6041	0.9194	0.8641	1.0784
5	0.9452	0.5481	0.8974	0.6245	1.4512
6	0.9345	0.6089	0.9142	0.7541	1.5412
7	0.8945	0.5412	0.8974	0.8412	1.5427
8	0.9421	0.5601	0.9215	0.7541	1.3245
9	0.9621	0.8941	0.8614	0.5412	2.5412
10	0.9124	0.8774	0.8164	0.7845	2.6542

along the horizontal direction. Hence, the horizontal thread distance (25) is doubled (51) and given as output by the proposed method.

Table 2 shows the five major GLCM features calculated for a set of 10 images taken from different fabrics of the same thread pattern with minor variation in orientation of camera and illumination. Images 4 and 6 are taken from the same fabric, Images 1, 3, and 7 are taken from the same fabric with different translations. Images 2 and 10 are taken from the same fabric with different illuminations. From Table 2, it could be observed that the feature values for Images 4 and 6 are close to each other. It is also evident that illumination difference causes a major difference in the features extracted.

Though the proposed method gives closer results for the parameter identification, the illumination variation affects the enhancement of thread, and hence the calculation of the parameters. The accuracy of the system can be improved by choosing appropriate camera parameters for thread visibility and controlling the environmental parameters such as uniform illumination.

5 Conclusion

The parameters for GLCM and Gabor vary based on their textural patterns of the fabric. Since it is difficult to predict the parameters for a particular fabric, wide range of these parameters are used by the automated systems for defect detection. Average response of these parameters is usually considered for decision-making, and hence influencing the accuracy of identification. The selection of correct parameters according to the fabric can give profitable computation cost and better accuracy in defect identification. This paper identifies these parameters by relating them to thread parameters. The thread parameters are identified from the Hough transform.

References

1. Yildiz1 K, Senyürek1 VY, Yildiz Z (2014) A new approach to the determination of warp-weft densities in textile fabrics by using an image processing technique. J Eng Fibers Fabr 2(1)
2. Priya S, kumar TA, Paul V (2011) A novel approach to fabric defect detection using digital image processing. In; 2011 international conference on signal processing, communication, computing and networking technologies (ICSCCN). IEEE
3. Tiwari Vikrant, Sharma Gaurav (2015) Automatic fabric fault detection using morphological operations on bit plane. Int J Comput Sci Netw Secur (IJCSNS) 15(10):30
4. Shi M, Fu R, Guo Y, Bai Sh, Xu B (2011) Fabric defect detection using local contrast deviations. In: Multimedia tools and applications, vol. 52, no. I, pp. 147–157
5. Lin Chun-Cheng, Yeh Cheng-Yu (2009) Texture defect detection system with image deflection compensation. WSEAS Trans Comput 8(9):1575–1586
6. Mohanaiah P, Sathyanarayana P, GuruKumar L (2013) Image texture feature extraction using GLCM approach. Int J Sci Res Publ 3(5):1
7. Pan, R et al (2010) Automatic inspection of woven fabric density of solid colour fabric density by the Hough transform. Fibres Text East Eur 18(4): 81
8. Yildiz K, et al (2014) A new approach to the determination of warp-weft densities in textile fabrics by using an image processing technique. J Eng Fabr Fibers (JEFF) 9(1)
9. Padmavathi S, Prem P, Praveenn D (2013) Locating fabric defects using gabor filters. Int Res Eng J Sci Technol (IJSRET) 2(8):472–478
10. Bovik A, Clark M, Geisler W (1990) Multichannel texture analysis using localised spatial filters. IEEE Truns PAMI 12(1):55–72
11. Michalewicz Z (1996) Genetic Algorithms + Data Structures = Evolution Programs, AI Series, 3rd edn. Springer, New York
12. Tsaia DM, Wua SK, Chen MC (2001) Optimal Gabor filter design for texture segmentation using stochastic optimization. Image Vis Comput 19:299–316
13. Daugman JG (1985) Uncertainty relations for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. J Opt Soc Am A 2:1160–1169
14. Sakhare K, Kulkami A, Kumbhakam M (2015) Spectral and spatial domain approach for fabric defect detection and classification. In: 2015 international conference on industrial instrumentation and control (ICIC), 28–30 May 2015

15. Hari CV et al (2009) Mid-point hough transform: a fast line detection method. In: India conference (INDICON), 2009 annual IEEE. IEEE
16. Wang Xin, Georganas Nicolas D, Petriu Emil M (2011) Fabric texture analysis using computer vision techniques. IEEE Trans Instrum Meas 60(1):44–56

Error Detection Technique for A Median Filter Using Denoising Algorithm

Maymoona Rahim and Ruksana Maitheen

Abstract This paper presents a modified design for error detection technique for a median filter, while processing the images using a digital processing system, acquisition stage capture impulsive noises along with the images. A classic filter used to eliminate. Some sliding windows of $n \times n$ matrices are created from the original image. The generated median is compared with the original pixel value at the center of the selected portion of the matrix corresponding to the image. If we found any error then the new median value is placed instead of the existing one. For storing the pixel values, a line buffer and register bank are used along with the median filter. Using the buffers the time taken for the median calculation is reduced, and as a result the total time taken for processing the image is also reduced.

1 Introduction

Image processing is a technique used to change an image into its digital form by performing some operations, to provide better visualization, image sharpening, image retrieval, and restoration. Using computers digital image can be manipulated by digital processing techniques. Image acquisition is the initial step in the image processing workflow because the process is carried out only when there is an image. Here, an image is retrieved from a hardware-based source into the system. The image which is retrieved is completely an unprocessed one.

M. Rahim (✉) · R. Maitheen
Department of ECE, Ilahia College of Engineering and Technology,
Affiliated to APJ Abdul Kalam Technological University (KTU), Muvattupuzha, India
e-mail: maymoonarahim77@gmail.com

R. Maitheen
e-mail: ruksanamaideen@gmail.com

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_119

There are some impulsive noises added along with the image. This is also known as salt and pepper noise. This is caused by sharp disturbances in the image. The image sensors acquire unwanted noises along with the captured image. So, it is necessary to remove the errors at the acquisition stage itself to reduce the chance of noise propagation. If the unwanted noises are not removed then it will propagate to the entire system and cause low-quality image with noise.

Median filter is a nonlinear filter used to eliminate the noises at the acquisition stage. It is also named as edge-preserving filter. It has varying traits with reference to linear filters. A nonlinear filter can generate output in a nonintuitive manner. For example, if the filter was based on five values as *Y0*, *Y1*, *Y2*, *Y3*, and *Y4* in median filter, initially, the values are sorted from lowest value to highest or vice versa. The value at the center is chosen as the median. That is, the value at position two is chosen as the median.

Field programmable gate arrays (FPGA) are semiconductor devices, which has a matrix of programmable logical blocks and are connected using programmable interconnects. FPGA can be reformed after manufacturing according to the design needs. FPGAs and targeted design technologies offer higher flexibility, quick time to market, and cheaper over all nonrecurring engineering costs for a wide range of video and imaging applications. In particular, SRAM-based FPGAs are easily and quickly reprogrammed. It provides high performance, high densities, and low cost.

The digital image sensors applicable in space applications are undergoing single event upsets (SEUs) due to the space radiations. It will damage the sensors or the entire image processing system. The cosmic rays or other energetic particles in the space collide with the endangered components of device, and may cause single event upsets. SEUs are the change of state caused by single ionizing particle striking the vulnerable part of the device. The SEUs are also called soft errors. SRAM-based FPGAs are also susceptible to the effect of radiation, so the implementation of median filter using SRAM-based FPGA can undergo SEUs which replaces the functionalities of median filter.

To protect the digital filters redundancy-based techniques such as dual modular redundancy (DMR) or reduced precision redundancy (RPR) is used. In DMR, the whole bit stream is replicated in case one should fail. Using these techniques error detection is possible in case of errors. In RPR, only the high-precision bits are replicated instead of replicating the full bits, which does not detect least significant change. This work is an enlargement of the paper error detection technique for a median filter presented in [1]. In this paper, the time taken for median calculation is reduced and as a result the time is consumed.

The paper is arranged as follows: Sect. 2 introduces the previous works related to the work. It gives the FPGA implementation of the original median filter. In Sect. 3, the proposed denoising architecture is explained. The experimental setup and simulation are reviewed in Sect. 4. Finally, Sect. 5 concludes the paper.

2 Related Works

The median filter is a nonlinear filter used to remove impulsive noises from the images. It is also used to provide care for protecting edges in an image. The filtering technique applied a two-dimensional window also called sliding window to the image and replaces the center pixel value with the median value calculated from the sliding window. Once the median replacement is done, the sliding window is shifted to another region of image and the process is continued until the whole image is processed. By these filtering techniques, it is possible to eliminate salt and pepper noises from the image and also preserve sharp edges. This is a type of intrinsic properties of nonlinear filter which is not achieved by a linear filter.

In sliding window filtering process, a 2D window is defined which is in an $N \times N$ square matrix (N is odd). N is odd value since to define exact center value for the window. N can be increased according to improve the filtering process and noise reduction. Thus, a 3×3 square window is chosen in this paper as the sliding window. Figure 1 explains the working of median filter using a 3×3 sliding window.

Initially, a sliding window of 3×3 matrixes is chosen from the original image. Now, the pixel values in the window are arranged in descending order. From the sorted values median is calculated. Since the order of the matrix is odd, it is easy to choose the mid-value. The new median value is compared with the value in the actual image. If any error occurs, then an erroneous signal is generated, and hence the actual center pixel value is displaced with the new median value calculated, i.e., the analyzed dead pixel is removed from the original image.

For the further processing of the entire image pixel, the sliding window of order 3×3 is shifted through the image. In FPGA system, this can be done with the help of a nine-pixel stream that eventually moves through the filter. Each group can be ordered using a node shown in Fig. 2 which consists of a comparator of eight bit and two 2:1 multiplexers. The two inputs are compared internally.

The classic sorting structure consists of 41 basic nodes shown in Fig. 3. Here, each box is identical exchange nodes illustrated in Fig. 2. The classic implementation has been substantiated to be far from the optimal one because the improved FPGA implementation has been developed over the years [2–5]. Using the Batcher bitonic sorter [3], a nine input sorting network can be created. This network consists of 28 nodes, which reduces the amount of FPGA used.

The architecture proposed by J. L. Smith uses the lesser number of exchange nodes. Here, the usage of FPGA is reduced since it accomplishes a nine-pixel partial sorting. Here, higher and lower pixel values are cluttered. It will not be a problem because it requires only the median value for replacement. Figure 4 illustrates this scheme.

Figure 5 depicts a fault-tolerant technique that activated an output error signal if a corrupted image pixel is detected. Then, a partial or complete reconfiguration can be performed to remove the error. Here, gray-shaded blocks are added to the original structure in Fig. 4 to create a range. Here, the time taken to create sliding window is larger. Once the sliding window is created by loading the nine-pixel values, the

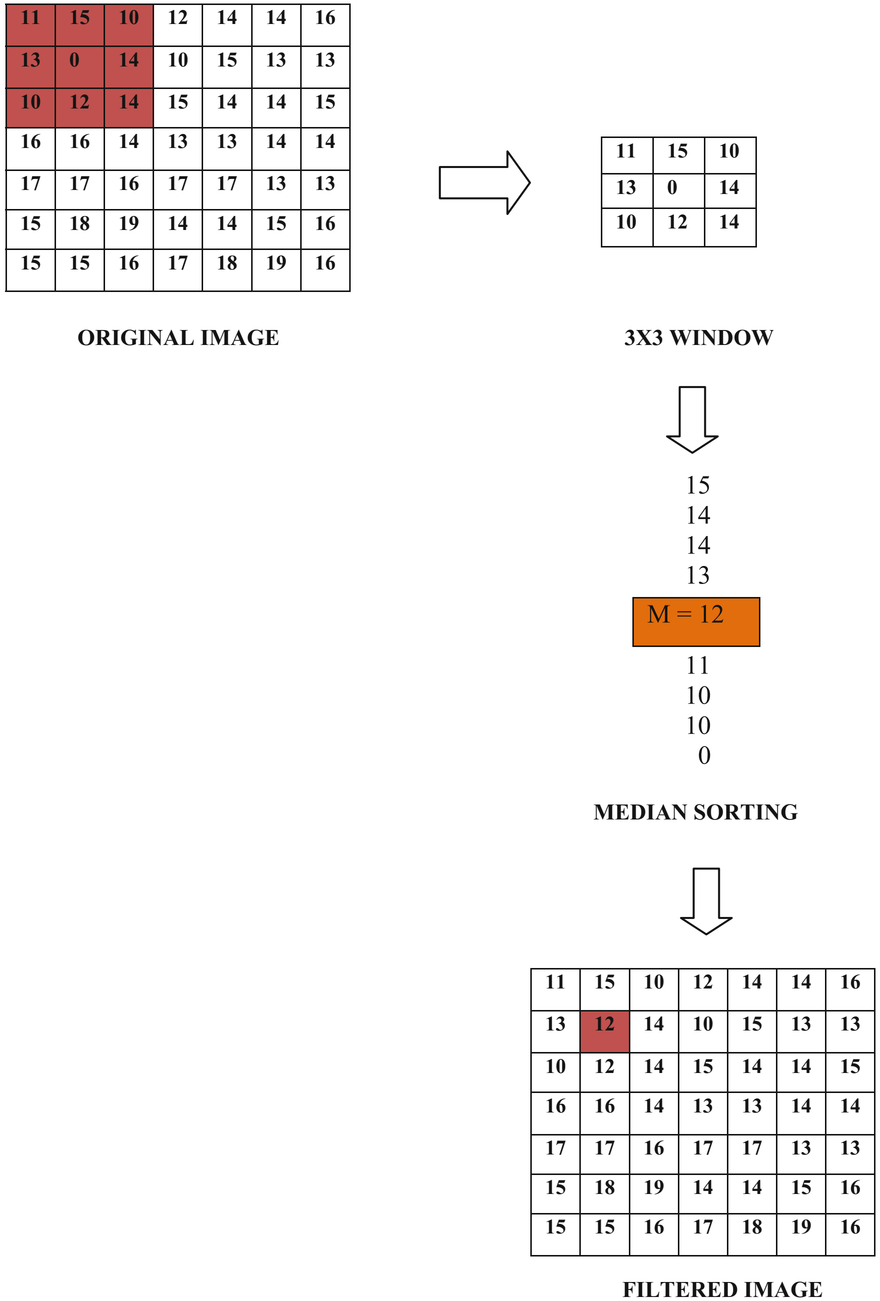


Fig. 1 Median substitution

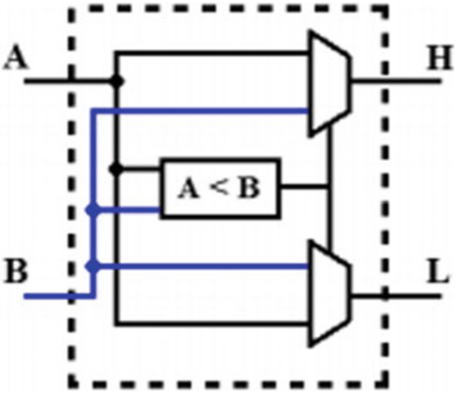


Fig. 2 Exchange node

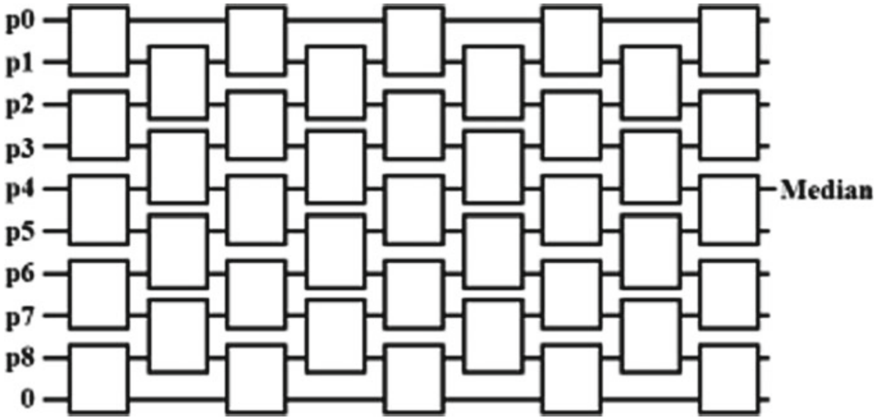


Fig. 3 Classic network

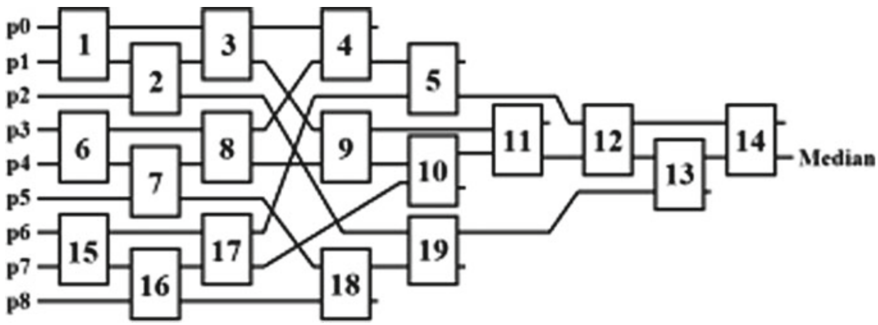


Fig. 4 Minimum exchange node network for nine pixels

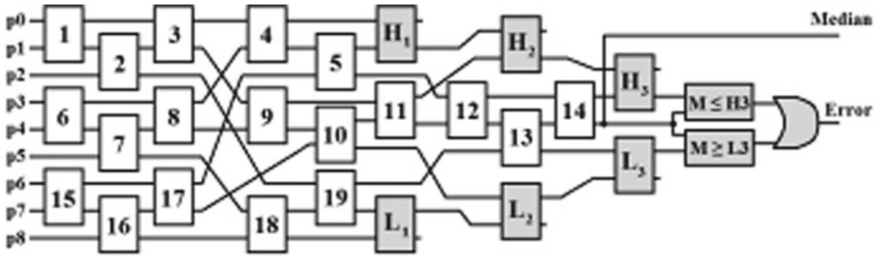


Fig. 5 Modified technique with reduced resources

median is calculated. If any error occurs, the center pixel value is reinstated with the new median value calculated. The process is performed again by shifting the sliding window and again loading the new nine-pixel values which make waste of time.

3 Proposed System

While using median filter for error detection in images the time taken to create sliding window each time is increased, and hence the entire image processing takes time. To avoid this problem, the proposed system is applicable. Instead of loading the pixel value every time, it is possible to use a line buffer for storing the pixel values, and hence the time can be consumed.

Buffers are physical memory allowed to store data temporarily and shift from one place to another if it is required. Generally, data stored in a buffer is retrieved from a hardware source. The output of buffers exactly matches the input. The system adopts a 3×3 mask, so it requires three scanning lines. Here, the concept of ping-pong arrangement is used. To recognize three scanning lines having two line buffers, it requires four crossover multiplexers. To store values in the odd rows, odd line buffers are designed and for values in the even rows even line buffers are designed, which is described in Fig. 6.

The register bank, contains nine registers, is used to collect the 3×3 pixel values of the current mask. Each register is connected serially to give the corresponding pixel values of a row and Reg4 keeps the median value of the denoised pixel value. Clearly, the denoising process for $p(i, j)$ does not start until $f(i+1, j+1)$ enters from the input device. The value which is stored in RB is used one by one by corresponding data detector and noise filter for the purpose of denoising.

Once the denoising process of the first set of pixel value is completed, the reconstructed value which is generated by the median filter is written back to the line buffers for storing the ith row for replacing the actual image. The denoising process is then shifted to the next set of pixel values, i.e., $p(i, j)$ to $p(i, j+1)$, it requires only three values they are ($fi - 1, j+2; fi$, j+2; $fi + 1, j+2$) and are required to read into the register banks. The register banks 2, 5, and 8 are used for this purpose, and the

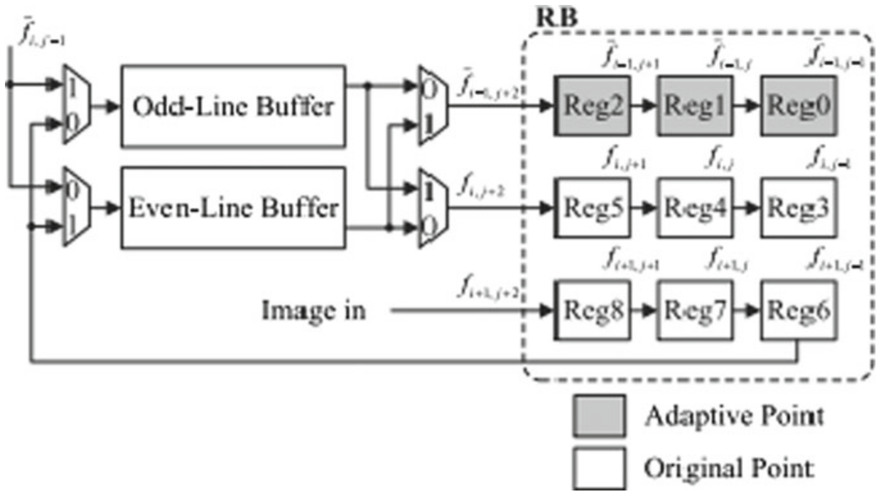


Fig. 6 Architecture of register banks

remaining six values are shifted to the corresponding register positions. Along with that the input value from the input source $f(i+1, j+1)$ is again written back to the buffer storing the previous rows for denoising process.

The multiplexers work in accordance with the select lines. The selection signals are set either one or zero for the purpose of denoising the odd row or even row, respectively. Figure 7 gives two examples. The figure illustrates the interconnection among line buffers and register banks. Assume, initially, we are going to denoise row 2, for that set all the four select line to zero, the samples written in row 1 and row 2 are stored in odd line buffer and even line buffer. The samples of the row are the inputs to the device, shown in Fig. 7a. Now, the previous value which is stored in register 6 and denoised output generated by the median filter are written to the odd and even line buffers. Now, the denoising process of row 2 is completed, the odd line buffer is filled with the samples of row 3. The denoised samples of row 2 are now stored in even line buffer. To denoise row 3, we have to select the select lines to 1. The previous value is stored in register 6. The denoised result from the median filter I against written back to any one of the line buffer which is demonstrated in Fig. 7b. On completion of the denoising process of row 3, even line buffer is filled with the samples of row 4 and odd line buffer is filled with the denoised samples of row 3.

4 Simulations and Experimental Results

The proposed solutions have been designed using Xilinx. In comparison with the previous technique, the sliding windows are generated with reduced time limit than the existing system which is presented in Fig. 8. Here, it takes some initial time delay

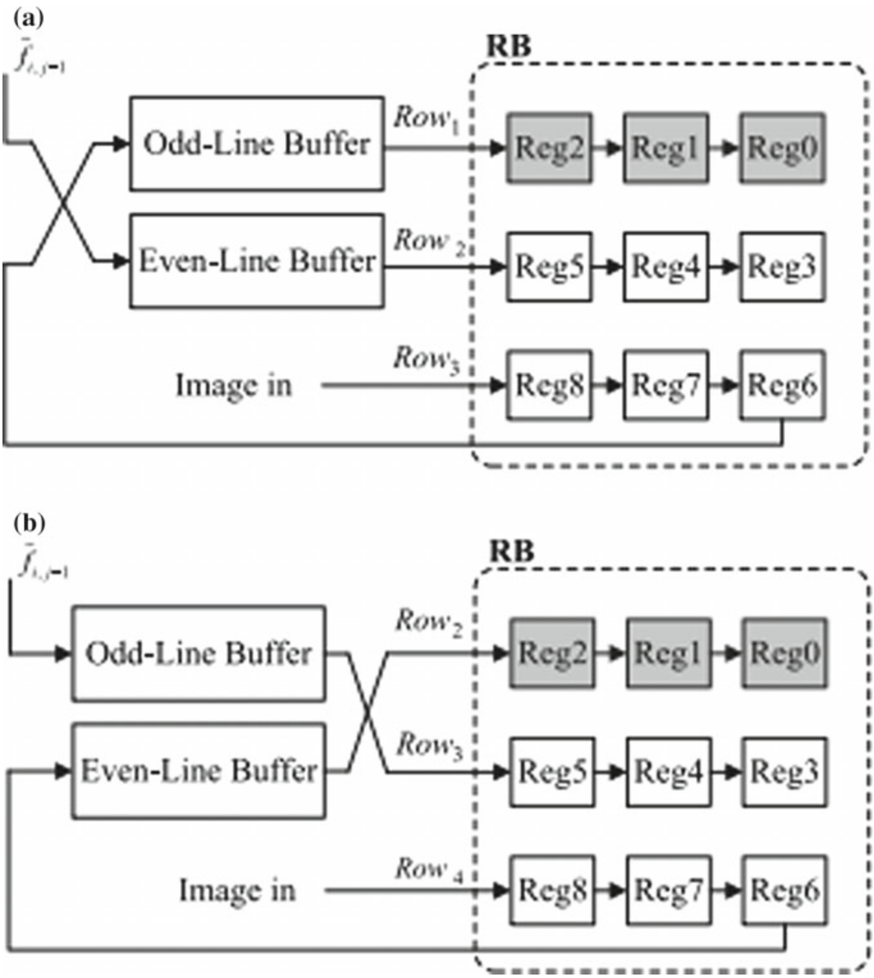


Fig. 7 Examples of the interconnections between two line buffers and RB

for loading the image into the buffer. After loading the image in each clock cycle, we are obtaining the median value from the sliding window without any delays. Hence, the image processing is carried out in a faster way than the existing techniques.

Table 1 gives the device utility during simulation. It uses lesser number of resources available which makes the system more advantageous.

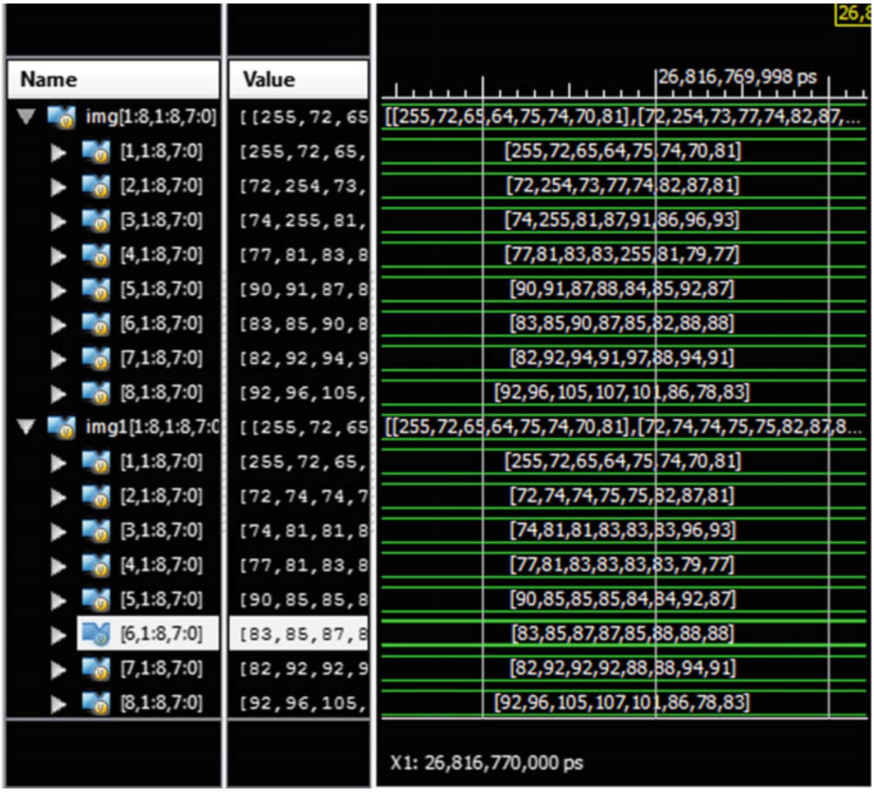


Fig. 8 Simulated results

Table 1 Device utilization summary

Device Utilization Summary (estimated values)			
Logic Utilization	Used	Available	Utilization
Number of Slice Registers	910	126800	0%
Number of Slice LUTs	3581	63400	5%
Number of fully used LUT-FF pairs	839	3652	22%
Number of bonded IOBs	18	210	8%
Number of BUFG/BUFGCTRLs	2	32	6%

5 Conclusion

Error detection during the initial stage of image processing is the most important part in image processing techniques. Median filters are helpful in removing the impulsive noises added along with the image and also these are edge-preserving filters. Along with the median filters, the use of line buffers and register banks reduces the time taken for image processing, and hence improves the system performance. The proposed system can be used in any image processing systems. This will improve the system performance by eliminating the noise at the initial stage itself.

References

1. Aranda LA, Reviriego P, Maestro JA. Error detection technique for a median filter
2. Fahmy SA, Cheung PYK, Luk W (2005) Novel FPGA-based implementation of median and weighted median filters for image processing. In: International conference on field programmable logic and applications, Tampere, Finland, pp. 142–147
3. Batcher KE (1968) Sorting networks and their applications. In: AFIPS spring joint computing conference, San Francisco, CA, pp. 307–314
4. Vasicek Z, Sekanina L (2008) Novel hardware implementation of adaptive median filters. In: 11th IEEE workshop on design and diagnostics of electronic circuits and systems (DDECS), Bratislava, Slovakia, pp. 1–6
5. Smith JL (1996) Implementing median filters in XC4000E FPGAs. Xcell 23:16

Retinal Image Processing and Classification Using Convolutional Neural Networks

Karuna Rajan and C. Sreejith

Abstract This study aims to develop a system to distinguish retinal disease from fundus images. Precise and programmed analysis of retinal images has been considered as an effective way for the determination of retinal diseases such as diabetic retinopathy, hypertension, arteriosclerosis, etc. In this work, we extracted different retinal features such as blood vessels, optic disc and lesions and then applied convolutional neural network based models for the detection of multiple retinal diseases with fundus photographs involved in structured analysis of the retina (STARE) database. Augmentation techniques like translations and rotations are done for expanding the number of images. The blood vessel extraction is done with the help of morphological operations like dilation and erosion and enhancement operations like CLAHE and AHE. The optic disc is localized by the methods such as opening, closing, Canny's edge detection and finally thresholding the image after filling the holes. The bright lesions (exudates) inside the retina are detected by the filtering operations and contrast enhancement after the removal of the optic disc. In this study, we experimented with different retinal features as input to convolutional neural networks for effective classification of retinal images.

1 Introduction

The retina is a thin layer of tissue lining the inner surface of the eyeball. The photoreceptor cells take light focused by the cornea and convert into neurosignals to the brain for visual recognition. Damages caused to the retina due to various diseases may gradually lead to loss of vision. Since the ageing of people across the world has been evolved as a major statistical trend, patients plagued by retinal diseases such as hypertensive retinopathy, age-related macular degeneration (AMD) and diabetic

K. Rajan · C. Sreejith (✉)
Calpine Labs, UVJ Technologies Pvt.Ltd, Kochi, Kerala, India
e-mail: sreejith.cherikkallil@calpinetech.com

K. Rajan
e-mail: karunarajan94@gmail.com

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_120

retinopathy (DMR) may increase in the future. Macular degeneration is a very common eye condition and a leading cause of vision loss among people Choi et al. [1]. Diabetic retinopathy is also a reason for visual disorder due to the damages of blood vessels in patients suffering from diabetes mellitus. Retinal diseases including retinal vessel occlusion, hypertensive retinopathy, Hollenhorst emboli and retinitis are also some of the prime causes of vision impairment. Specifically, it is all because of our lifestyle changes and unhealthy practices in our day-to-day life. In many cases, visual loss can be avoided if early detection and timely treatment are provided prior to the initial stage of blindness. Thus, more precise screening programmes aided with modern approaches are needed for early treatment so as to reduce the socio-economic hardships of visual loss caused by retinal diseases. The diabetic retinopathy screening that uses fundus photographs has been widely adopted for diabetes patients Abrmoff [2]. Conducting AMD screening and DMR screening is cost-effective in a public health scenario. However, manual analysis of multiple fundus photographs for an accurate result requires efforts and expertise of the ophthalmologist. Thus, comes the relevance of the retinal image processing. It deals with the analysis of the retinal features that may help us in the diagnosis of this kind of diseases.

The entire paper is devoted to retinal image analysis and classification methods and their clinical implications. Blood vessels, optical disc, optic nerve, lesions and fovea are some of the most important structures of the human retina and are mostly used for several applications such as pathology detection inside the retina. Detection of these important structures manually is time consuming and depends upon the expertise of the user. The proposed work aims at developing an automated system that can predict the diseases by analysing these features from the retinal fundus images. The fundus (the bottom or base part of an organ) is the part of the inner eye that can be seen during an eye examination while looking through the pupil. The inner lining of the eyeball, including the retina (the light-sensitive layer), optic disc (the brightest spot inside the eye and also the head of the nerve to the eye) and the macula (the small spot in the retina where vision is the sharpest) together constitute the retinal fundus. This work focuses only on the retina; nevertheless, a brief review of the eye anatomy can be found from the following Fig. 1 [3]. In this study, we applied deep CNN for fundus photography analysis in multi-categorical disease classification. This paper explains the retinal image processing and disease diagnosis on an open retinal image database (STARE). The proposed work has two stages: feature Analysis and classification.

2 Literature Survey

The problem of retinal image analysis has already received unique attention and many works were published on this topic. Sinthanayothina et al. [4] detected features such as fovea, optic disc and blood vessels in their work. More than 100 retinal images were preprocessed by means of adaptive, contrast and local enhancement. Those regions with the largest variation in the intensity of neighbouring pixels were identified

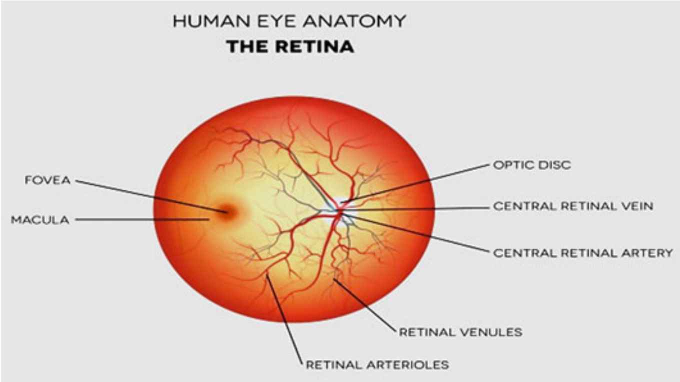


Fig. 1 Human eye anatomy

and thus optic discs were localized. Blood vessels were recognized by means of a perceptron. The foveae were identified using correlation matching. Hoover et al. [5] proposed a novel algorithm called fuzzy convergence. This algorithm identified the optic nerve as the converging point of the blood vessel network. In the absence of a strong convergence, the method identifies the optic nerve as the brightest region in the image after contrast equalization. Snchez et al. [6] proposed the Fisher linear discriminant analysis for the detection of hard exudates in the retinal images. Hard exudates were identified by making use of the colour features to define the feature space. So as to select one of these models they calculated a quantitative metric (J) to evaluate the performance of different colour spaces. Mishra et al. [7] proposed the methods like cup-to-disc ratio (CDR) and inferior superior to nasal temporal side) (ISTN) for the aim of glaucoma detection. In this work, active contour models were utilized for the identification of cup and optic disc. Veins were separated by morphological strategies as it helps in smoothing the picture.

Aslam et al. [8] proposed methods such as PCA and Hough transforms for the localization of optic nerve head. But it was not a success for fundus images having a large number of white lesions. For the fovea detection, the proposed method is the HSI transformation. The vascular segmentation was done with the help of techniques like matched filters, vessel tracking, neural net and morphological processing. Fundus photographs were used for the entire work. [9, 10] proposed a template matching method for the detection of blood vessels. For applying the entire algorithm of the matched filter, the G-plane of the image was considered. After enhancing the contrast of the image, the median filter was used to remove the noise. A bank of filters (matched filters) was applied to the image to localize and segment the veins. In this, an advanced, non-parametric tree-type classifier random forests (RF) was used for classification.

Tjandrasa et al. [11] proposed the Hough transform and active contour models that implemented the detection of the optic nerve head in retinal fundus images. The procedure begins with the image improvement utilizing homomorphic filtering for brightness correction, then continues with the expulsion of veins from the images. Blood vessels were detected by the method thresholding. Nguyen et al. [12] proposed the basic line detector which was first used as a means for vessel-background classification. Supervised methods like Kalman filter and pixel classification approaches are also used. Top hat filter, Gabor filter and matched filter were some of the techniques for the blood vessel extraction. Radha et al. [13] proposed methods like plane separation, contrast enhancement and morphological process for the blood vessel extraction. Exudates were detected and shown accurately with the help of K-means clustering and the extracted features were trained with the probabilistic neural network successfully [14, 15]. Melinsak et al. [16] suggested max-pooling layered convolutional neural networks (MPCNNs) for the retinal vessel segmentation.

Chandore and Asati [17] also proposed convolutional neural networks for the detection of diabetic retinopathy. The features discussed in this paper for the detection of diabetic retinopathy were optic disc, macula, exudates, microaneurysms, haemorrhages, etc. They simplified the labels to only two classes (without DR-class 0, with DR-class 1). On this simplified model and dataset, they got the precision of about 0.81 for class 0 and 0.88 for class 1. In the paper, Tan et al. [18] proposed some solutions not just to locate but to simultaneously segment vasculature, optic disc and fovea. They used a seven-layer convolutional neural network (CNN) to classify every pixel into one of the four classes: background, blood vessels, optic disc and fovea. They used a seven-layer CNN together with random forest algorithm to extract retinal vasculature. However, in their work, CNN was used only as a trainable feature extractor, the classification was in fact performed by an ensemble of random forest. Dasgupta et al. [19] proposed a fully convolutional neural network architecture for blood vessel segmentation.

Choi et al. [1] proposed deep learning approaches for the multi-class disease classification. In this approach, they have utilized the transfer learning along with the traditional classifiers like the random forest. They have used the VGG19-TL-RF model for the classification purpose. In this study, they obtained results with an accuracy of 30.5%, relative classifier information (RCI) of 0.052 and Cohen's kappa of 0.224 when all 10 categories were included. When the concept of transfer learning was added along with the ensemble classifiers they got an accuracy of 36.7%, 0.053 RCI and 0.225 kappa for the same 10 retinal diseases classification problem.

The purpose of this research is to develop an automated system for the analysis of retinal images, and thereby the detection of disease from 10 categories. Hence, the proposed design is a novel approach that combines the conventional analysis methods along with the modern deep learning techniques and thus leads to the detection of the retinal diseases as early as possible.

3 System Design

This section explains how the entire framework is designed and its working. The proposed system design is depicted in Fig. 2.

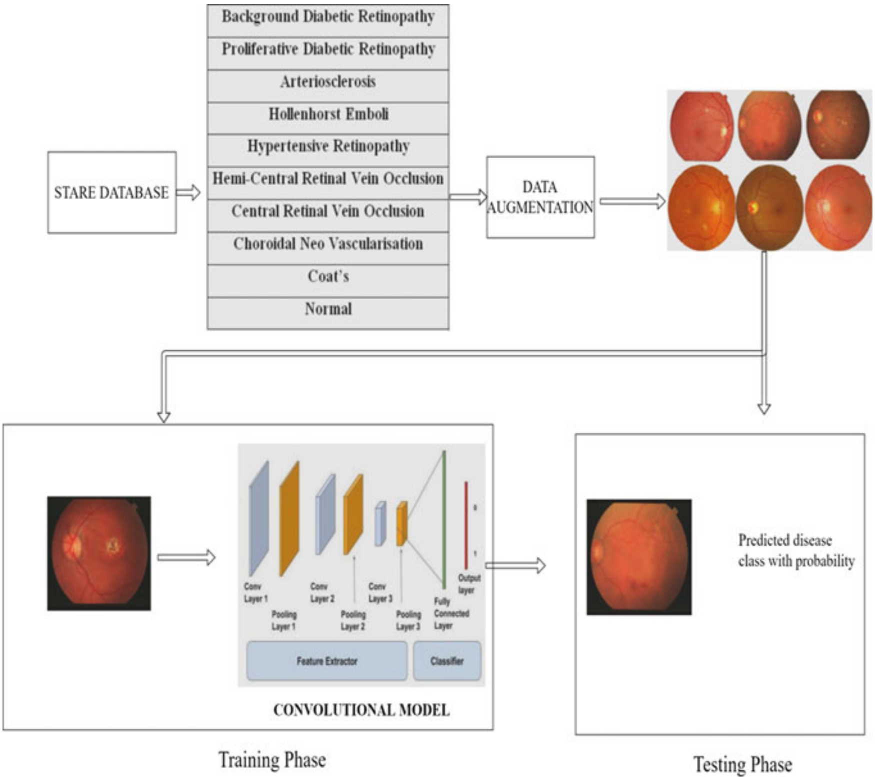


Fig. 2 Proposed system design

3.1 Dataset

For this study, the publicly accessible retinal image database for structured analysis of the retina (STARE) [20] was used, so as to develop a multi-categorical disease detection system with the help of deep learning approaches. The database contains retinal colour images captured with the help of a TRV-50 fundus camera at an angle of 35 degrees at a resolution of 605 × 700 pixels. It contains 397 images in 14 disease categories. The major problem with this dataset is unequal categorization. For example, the dataset includes only a single image of cilioretinal artery occlusion, while more than 60 images were included in the BDR. The dataset was balanced and cleaned by removing redundant data and eliminated those classes with least amount of data. Data augmentation techniques such as translation, rotation and brightness change were used for expanding the dataset. We selected 10 classes, created 100 fundus images per each disease category and eliminated redundant images. Specifically, we obtained samples with translation from the range [–10%, +10%] of the image width, with the rotation of [90, 45] and with brightness change from a range of [-10%, +10%].

3.2 Feature Extraction

Retinal feature localization and extraction are some of the important steps involved in the initial stages of image processing. Various image processing methods were used to extract/localize the prime features inside retina that may help us in further stages of disease detection. In this work, we used morphological operations for the extraction and localization of retinal features such as blood vessels, optic disc and yellow deposits. Blood vessels can be extracted with the help of a number of morphological operations such as opening (erosion followed by dilation), closing (dilation followed by erosion), contrast limited adaptive histogram equalization, adaptive histogram equalization, etc. The whole processing is done on the extracted green channel.

Yellow deposits/exudates (soft or hard) are because of the leakage of blood from broken blood vessels due to diabetes, hypertensive retinopathy, neuroretinitis and retinal vein occlusion. This kind of exudates or lesions that cause vision-threatening disorders can be detected by performing the Gaussian blurring on the retinal image and then applying binary thresholding (threshold=140) after converting it to grayscale.

Optic disc is an anatomical component of the retinal image which behaves as a recognizable proof for the retinal disease detection. It is the brightest spot inside the retina. The optic disc can be detected with the assistance of a series of morphological operations, such as closing (with a kernel size of 5×5), then erosion operation is applied with two iterations, dilation is then performed with three iterations and after that gradient operation is performed. Canny's edge detection algorithm is then applied to the image and a closing operation towards the end, to locate the optic disc.

3.3 Classification

The classification stage involves two stages, namely, training stage and testing stage. The training stage involves identifying the representative classes and developing a numerical description of the attributes of each class type through training set. Later, in the classification stage, the dataset is classified to the class it resembles the most, and finally the output stage generates the disease category of the corresponding image along with the probability. In a convolutional layer, all neurons apply the operation called convolution to the inputs, and hence they are also known as convolutional neurons. All the neurons are fully connected to the neurons in previous and post-layers. It is usually made of filters and feature maps. A pooling layer will be used just after each convolutional layer so as to decrease the spatial size (only width and height, not depth). This reduces the number of parameters, and hence computation complexity can be reduced. Also, less number of parameters avoid the overfitting. Each neuron in each layer in a fully connected layer will be connected to each other. A normal flat feedforward neural network layer is used at the end of the network after the extraction of essential features. In this work, we are using the 11-layer convolutional model with three convolutional layers, three max-pooling layers, one dropout layer, one flatten layer and three dense layers. The classification of images on the basis of disease categories are done with the help of these layers in the CNN. Different models were built on the basis of channels and extracted features

4 Experiments and Results

This section describes the techniques used in the preprocessing, the feature extraction and the detailed explanation of the models built for the purpose of classification. In this work, top 10 best classes were selected and after the data augmentation, the dataset was equalized among the classes with 100 images each. Important features needed for the disease diagnosis are extracted and localized in the image processing stage. Convolutional models are built on the basis of this data and diagnosis of the disease also occurs at this stage.

4.1 Image Preprocessing and Feature Analysis

In the feature analysis stage, important retinal features were localized and extracted needed for the disease detection. The process of vein extraction was done after all the image preprocessing tasks mentioned in the underlying stage. Morphological operations were adopted for this stage. The optic disc and lesions were also localized

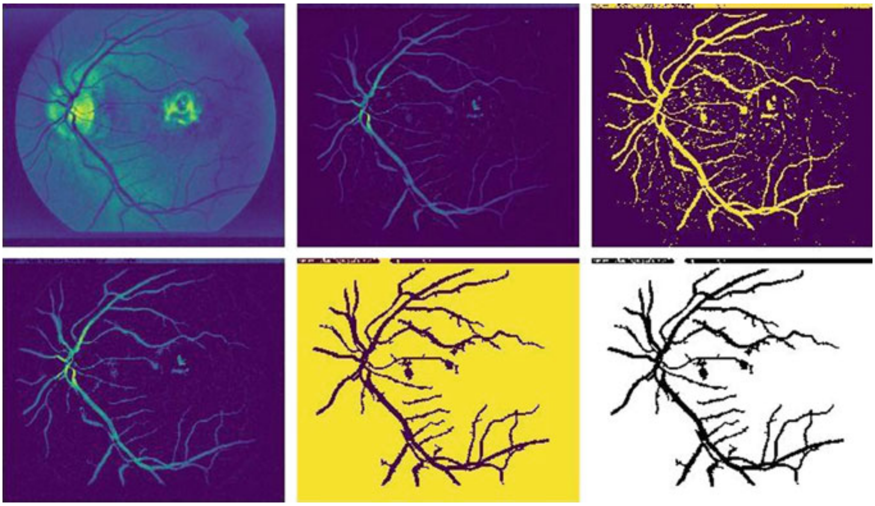


Fig. 3 Stages of blood vessel extraction

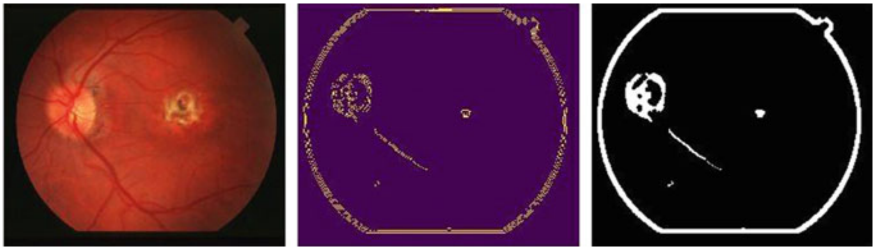


Fig. 4 Optic disc localization

Fig. 5 Lesion localization

with the help of some enhancement and edge detection operations along with the above-mentioned morphological operations (Figs. 3, 4 and 5).

4.2 Classification

Table 1 shows the accuracy of different CNN models with different retinal features as input.

In the stage of classification, we make use of the convolutional models that are made up of three kinds of layers, convolutional (calculating the output of the connected local input neurons studying the patterns), max pooling (subsampling the inputs) and fully connected layers (predicting the class by allocating the scores). The

Table 1 Results for the classification stage

Model name	Training accuracy (%)	Test accuracy (%)
RGB model	97.8	42
Green channel	95	35
Greyscale	93	32
Blood vessel	87	38

10 class classification problem was evaluated on the basis of different models tried on the different channels and extracted features.

5 Conclusion and Future Scope

This work is a primary attempted to construct deep learning models for multi-categorical classification problems in order to detect multiple retinal diseases. The deep learning techniques in this study were promising to be applied for the diagnosis. The problems we faced are low quality and quantity of the data, variations in the shape of the structures and the ambiguous boundaries of the structures that to be extracted. As a future work, transfer learning can be adopted to improve the classification performance in order to detect multi-categorical retinal diseases. Also, this work can be modified by collecting more data in various disease categories and experimenting with deeper models that combine various retinal features.

Acknowledgements This work was fully funded by Calpine Labs, UVJ Technologies, Kochi, India. We are also immensely grateful to Mr. Bijeesh Devassy, Project Manager, UVJ Technologies and Dr. Asharaf S, Associate Professor, IIITM-K for sharing their pearls of wisdom with us during the course of this research.

References

1. Choi JY, Yoo TK, Seo JG, Kwak J, Um TT, Rim TH (2017) Multi-categorical deep learning neural network to classify retinal images: a pilot study employing small database. PLoS ONE 12(11):e0187336
2. Abràmoff MD, Garvin MK, Sonka M (2010) Retinal imaging and image analysis. IEEE Rev Biomed Eng 3:169–208
3. https://www.virginiaeyeconsultants.com/procedures/eye-conditions/retina/
4. Sinthanayothin C, Boyce JF, Cook HL, Williamson TH (1999) Automated localisation of the optic disc, fovea, and retinal blood vessels from digital colour fundus images. Br J Ophthalmol 83(8):902–910
5. Hoover A, Goldbaum M (2003) Locating the optic nerve in a retinal image using the fuzzy convergence of the blood vessels. IEEE Trans Med Imag 22(8):951–958
6. Sánchez CI, Hornero R, Lopez MI, Poza J (2004) Retinal image analysis to detect and quantify lesions associated with diabetic retinopathy. In: 26thAnnual International Conference of the IEEE engineering in medicine and biology society, IEMBS'04, vol 1. IEEE, pp 1624–1627
7. Mishra M, Nath MK, Dandapat S (2011) Glaucoma detection from color fundus images. Int J Comput Commun Technol (IJCCT) 2(6):7–10
8. Patton N, Aslam TM, MacGillivray T, Deary IJ, Dhillon B, Eikelboom RH, Yogesan K, Constable IJ (2006) Retinal image analysis: concepts, applications and potential. Prog Retinal Eye Res 25(1):99–127
9. Verma K, Deep P, Ramakrishnan A (2011) Detection and classification of diabetic retinopathy using retinal images. In: 2011 Annual IEEE India conference (INDICON), pp 1–6. IEEE
10. Priyadharsini BH Devi MR (2014) Analysis of retinal blood vessels using image processing techniques. In: 2014 international conference on intelligent computing applications (ICICA), pp. 244–248. IEEE

11. Tjandrasa H, Wijayanti A, Suciati N (2012) Optic nerve head segmentation using hough transform and active contours. Indones J Electr Eng Comput Sci 10(3):531–536
12. Nguyen UT, Bhuiyan A, Park LA, Ramamohanarao K (2013) An effective retinal blood vessel segmentation method using multi-scale line detection. Pattern Recogn 46(3):703–715
13. Radha R, Lakshman B (2013) Retinal image analysis using morphological process and clustering technique. Signal Image Process 4(6):55
14. Wang H, Hsu W, Goh KG, Lee ML (2000) An effective approach to detect lesions in color retinal images. In: 2000 Proceedings of IEEE conference on computer vision and pattern recognition, vol 2. pp. 181–186, IEEE
15. GeethaRamani R, Balasubramanian L (2016) Retinal blood vessel segmentation employing image processing and data mining techniques for computerized retinal image analysis. Biocybern Biomed Eng 36(1):102–118
16. Melinščak M, Prentašić P, Lončarić S (2015) Retinal vessel segmentation using deep neural networks. In: VISAPP
17. Chandore V, Asati S (2017) Automatic detection of diabetic retinopathy using deep convolutional neural network
18. Tan JH, Acharya UR, Bhandary SV, Chua KC, Sivaprasad S (2017) Segmentation of optic disc, fovea and retinal vasculature using a single convolutional neural network. J Comput Sci 20:70–79
19. Dasgupta A, Singh S (2017) A fully convolutional neural network based structured prediction approach towards the retinal vessel segmentation. In: 14th International symposium on biomedical imaging, pp. 248–251. IEEE
20. STARE Dataset, http://cecas.clemson.edu/~ahoover/stare/

Comparison of Thermography and 3D Mammography Screening and Classification Techniques for Breast Cancer

Sureshkumar Krithika, K. Suriya, R. Karthika and S. Priyadharshini

Abstract Breast cancer, without doubt is one of the leading reasons for fatality among women in the world after lung cancer. Awareness and accessibility to better screening and treatment protocols will have a major impact in improving the survival rates. Moving away from the traditional methods of mammography and biopsy methods, newer techniques provide faster and efficient results to ensure early start of treatment. Therefore, a comparison study has been performed to weigh the pros and cons of thermography and 3D mammography as screening methods, followed by their respective processing and classification procedures. The ease of screening, extent of radiation, percentage of false positives, efficient segmentation, clustering, and novel classification are all considered and a conclusive result is obtained determining the better of the two processes. This could potentially revolutionize the way breast cancer is diagnosed and treated for women of all ages and walks of life.

1 Introduction

In the era of automation that we live in today, it is vital to devise methods that take down the barrier of manual detection, processing, and classification of medical images. This will not only save time but also be instrumental in reducing errors that, in the case of cancer, may cause unimaginable emotional and physical strain to the patients. Minimal access to advanced medicine and poor awareness in the developments of the medical and scientific fields are the main reasons for the less survival rates in breast cancer, especially among developing countries. The staggering numbers in various countries are Gambia (12%), Algeria (38.8%), India (52%), and Brazil (58.4%) and this situation can be ameliorated by sensitizing the use of modern medical practices. In light of the issue, the various screening processes are compared and their respective advantages and disadvantages are concluded. Among the different processes, the respective processing and segmentation steps are carried

S. Krithika (✉) · K. Suriya · R. Karthika · S. Priyadharshini
Easwari Engineering College, Chennai, India
e-mail: krithi96@gmail.com

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_121

out for thermogram and 3D mammogram [1]. Since they are of different nature, that is, thermogram is an RGB image and mammogram is a black and white image, naturally their processing steps differ and cater to their specific needs, while mammogram is filtered and processed by wavelet and discrete cosine transforms. Thermogram is processed by curvelet transform. Once the required features are extracted from the image, a database is created and used for classification in SVM which is a machine learning algorithm that could provide automated results by comparing the acquired features with the existing support vector values. The SVM classifier in place of biopsy can tremendously reduce the 2–3 weeks waiting period for results giving a clear picture of the manifestation of the unwanted growth.

2 Overview of Various Screening Methods

Breast self-exam is very instrumental in discerning a problem at the early stage. To confirm the actual presence and manifestation of the tumor, rigorous screening methods are being used around the world. Apart from mammogram and thermogram, the most prevalent ones are as follows:

- Sonogram,
- MRI imaging, and
- Molecular imaging.

As per traditional practice, these methods are followed up biopsy. During the wait time, there is a high possibility for the development of contralateral breast cancer (CBC) [2]. This is a condition where there is growth of cancer on the other breast while the original one is present or it develops post-removal of earlier cancer.

2.1 Sonogram

An ultrasonic wave is sent through the breast tissue and the presence of any abnormality will present itself as reflections from the surface that is captured as a sonogram. This is used to check whether a lump is filled with fluid (a cyst) or if it is a solid lump. This method can only be used as an auxiliary to mammogram and not as an independent procedure. The possibility of false positives is also comparatively high to that of mammogram (Fig. 1).

2.2 MRI Imaging

Breast magnetic resonance imaging (MRI) is more invasive than mammography because a contrast agent is given through an IV before the procedure. Magnetic

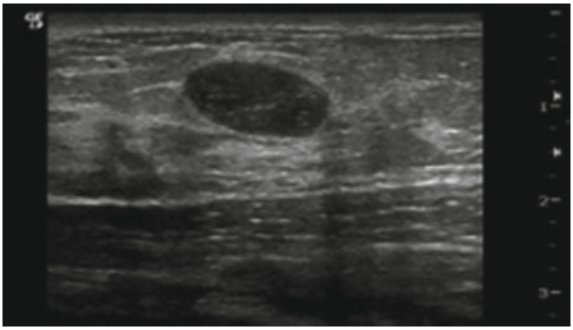


Fig. 1 Sonogram depicting a fluid cyst (*Source* https://www.dic-kc.com/blog/2016/mammogram-reports-and-bi-rads-categories-4-5)

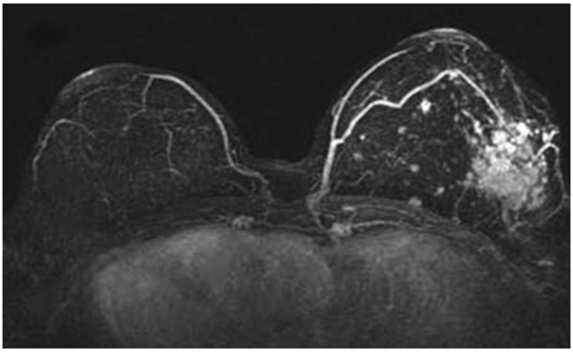


Fig. 2 MRI imaging of cancer in the right breast (*Source* https://breast-cancer.ca/mrifacts/)

fields are used to create images of the breast which in turn aid in staging of the cancer more than detection. Though accuracy levels have improved in recent times, this method was not opted for a very long period due to its complexity and prolonged time of operation (Fig. 2).

2.3 Molecular Imaging

Molecular imaging is a new and upcoming method. Though it may give the required results with the advent of new technology and safer equipment, this procedure is still uncommon in developing countries. In addition to that, a nuclear tracer is injected into the veins to clearly illuminate the structure of the tumor [3]. This may cause long-term defects in the patient. Since this mechanism depends on the breast structure and tissue it is advisable only for older women with denser breasts (Fig. 3).

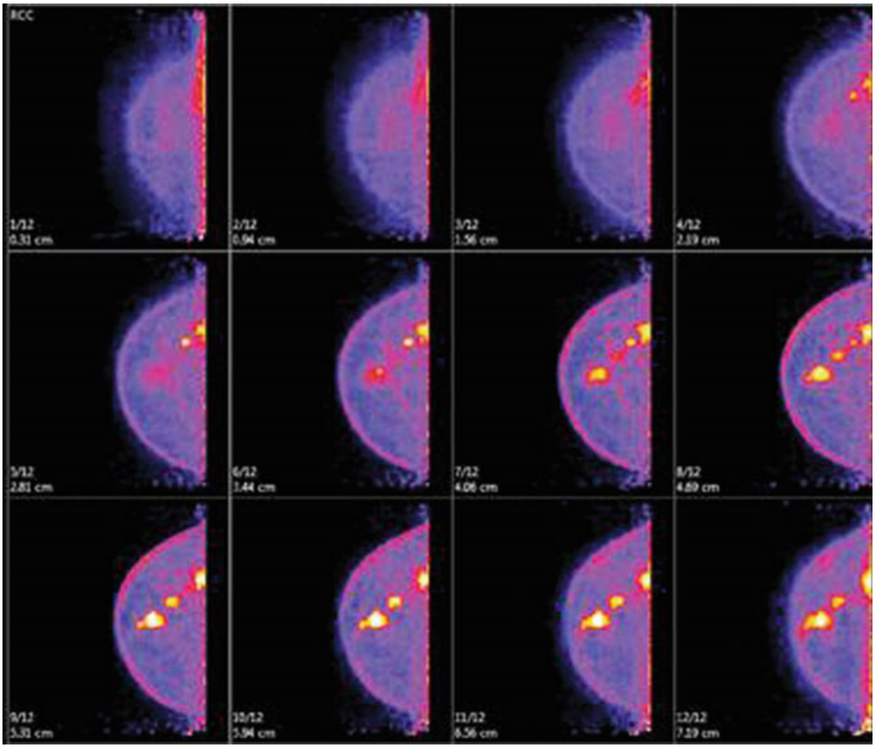


Fig. 3 Molecular imaging (*Source* HHS Author Manuscripts PMC2748346, p. 1073)

3 3D Mammography

3D mammography is similar to 2D mammography in the fact that it uses X-rays to produce images of breast tissue in order to detect lumps, tumors, or other abnormalities. Though it uses the same radiation, the advantage from traditional mammography is that it produces high-quality images, producing a crystal clear 3D construction of the breast, enabling the doctor to find out if there is anything to be concerned about. It captures multiple slices of the breast at various angles. 3D mammograms have the ability to discern a cancer cell from within layers of tissue as supposed to the 2D mammogram that only shows the growth of a mass as shown in Fig. 4.

3.1 Advantages and Disadvantages

Research has shown that mammograms are the reason for the decline in fatality rates in women all over the globe. Despite this feedback, definite answers can be obtained

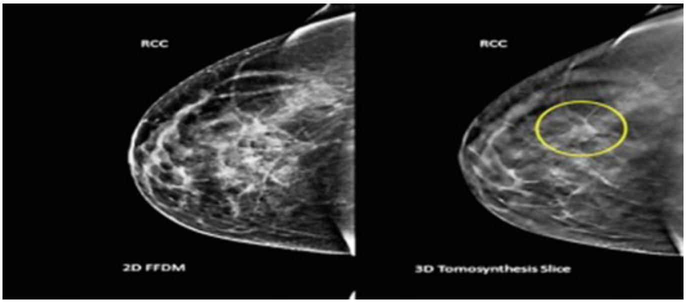


Fig. 4 Comaprison of 2D and 3D mammograms (*Source* www.google.com)

only after multiple screenings which expose the patient to constant and recurrent radiation [3]. This might set off the growth of more cancerous cells. Research shows that there is multiple recall screening sessions that are scheduled due to initial false reports. The 3D mammogram has a false-positive rate of up to 4%, though it is small number, it has the ability to cause mental and emotional trauma until the discrepancy is cleared.

3.2 Acquisition Procedure

The woman is positioned before a 3D mammography machine and her breasts are compressed by two metal plates to hold the flesh in place and still for the duration of the screening which usually lasts about 10–20 min. When ready, the machine is started and a robotic arm will move in an arc over the woman's breasts while multiple X-ray images are taken in various angles to reconstruct the 3D image of the breast as best as can be done. The dose of radiation is similar to 2D mammography but slightly higher in quantity to produce images with better contrast and structure.

3.3 Preprocessing

After the image is acquired, noise filtering is carried out using Gaussian or Wiener filter. Adaptive filters like Wiener are better for removing nonlinear noise. The image is first contrast enhanced and then salt and pepper or Gaussian noise is added and removed using adaptive filtering [4] (Fig. 5).

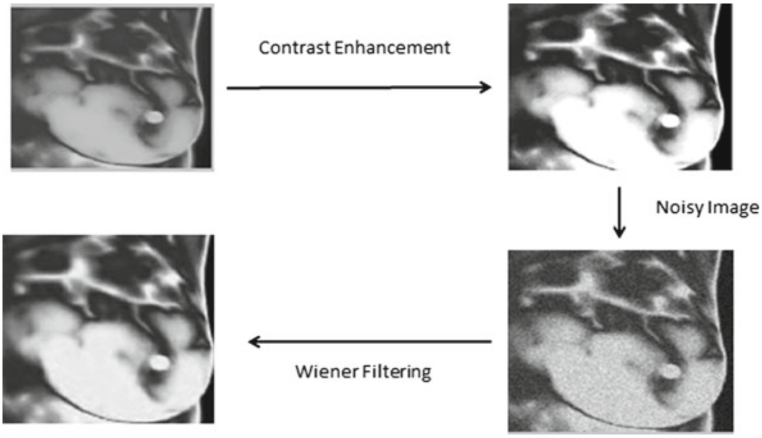


Fig. 5 Noise filtering of the 3D mammogram (*Source* Results from this study)

3.4 Feature Extraction

Extracting necessary features from the image is the main process before starting the classification process. Here, instead of discrete cosine transform, we use spherical wavelet transform. The spherical wavelet transform (SWT) proposed by Starck. It has better results as far as discerning features and evaluating them. SWT was initially used only for astronomical images but due to its focus on minute discontinuities and its inherent property to visualize the image in many parts to uncover all the necessary features, it has been deemed suitable for mammograms [5]. The main goal of this method is to reduce redundancy which results in smaller datasets and easier computation for classification. It is an isotropic transform that utilizes a wavelet pyramid to capture statistical features in an isotropic field. This helps in extracting structural features in medical images (Fig. 6).

Once the features are extracted using SWT, it is stored in a testing database ready to be compared with the training database using the SVM classifier.

4 Thermography

Thermogram is the depiction of infrared radiations from objects which exist at temperatures above absolute zero. The intensity of radiation increases with temperature and is portrayed as various temperature levels on a thermal image [6]. A thermogram is captured in a private room where the patient's body is cooled to a very low temperature causing a drop in normal blood flow. A thermo cam captures the radiations from various parts of the body, in this case the breasts. The presence of a cancerous growth is associated with the excessive formation of blood vessels and inflammation

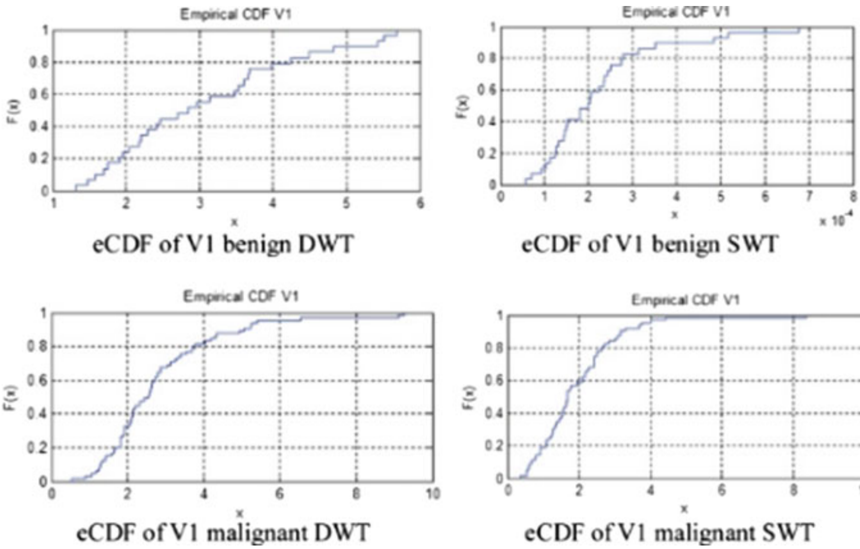


Fig. 6 Comparison of DWT and SWT for benign and malignant cells (*Source* Technol Cancer Res Treat, p. 507)

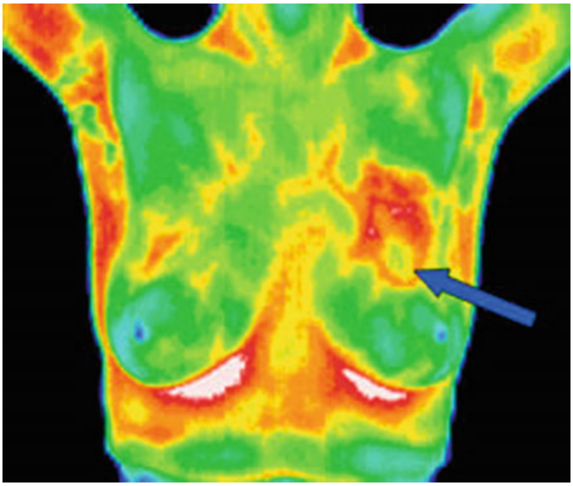


Fig. 7 Thermogram showing a high-temperature region with tumor growth (*Source* http://www.integratedhealthsolutions.com/medical-thermography/breast-cancer-risk/)

in the breast tissue. These show up on the thermal image as areas with a higher skin temperature. On further inspection of the breast, the extent of the tumor spread and the layers it has penetrated can be found out (Fig. 7).

4.1 Advantages Over Other Screening Methods

Thermography, though is a relatively new approach has a lot of advantages over the other methods, thereby making it highly sought after procedure. Researches have shown that thermography has the ability to diagnose breast cancer at least 10 years in advance [3]. Its most distinguishing benefit is that it can be used for detecting tumors in women below the age of 40 which is not the case for other procedures. It does not depend upon the structural aspects of the breast but the thermal aspect of the tissue, and hence it gives more conclusive answers for all women irrespective of their age, size of breast, tissue density, or whether they are pregnant and lactating. Since no radiation is used, no chemical inputs into the bloodstream and any other invasive technique, thermograms are an easy and affordable way to screen for abnormality. Repeat scanning is an issue in mammograms. This exposes the patients to excess radiation, whereas thermogram is a one-time procedure. Due to the high probability of real-time imaging, thermograms are used to monitor patients even after surgery for removal of infected tissue. The RGB infrared image is better suited for processing and feature extraction making the classification process simpler. Though it is effective and 100% harmless, obtaining high-contrast images with less signal-to-noise ratio is still proving to be difficult. This can be corrected later on using transforms and preprocessing adaptive filters. Despite these minor drawbacks, thermography is considered the best choice for screening.

4.2 Processing and Segmentation

After acquisition of the image, a set of processing steps that carry out de-noising and enhancement operations on the image is carried out. Other screening methods produce black and white images as supposed to the RGB of the infrared image, and therefore the processing steps are different and in various stages. Primitive noise cancellation filters are insufficient and adaptive filters like Gaussian filters are used to remove the nonlinear noise. After filtering, contrast enhancement is done on the image to sharpen the edges and intensify the difference in the background and foreground colors giving specificity to the colors in the image. Areas that exhibit as red-colored portions generally may host abnormal cells and the blue-colored regions, on the contrary may be devoid of any tumor (Fig. 8).

Preprocessing is followed by segmentation where the region of interest (ROI) is deducted. This is vital in understanding the extent of the growth and to make targeted treatment plans. For the acquired contrast-enhanced image, a segmentation process called clustering is performed. Here, the image is segregated and sequestered based on the K-means algorithm. It separates the initial image into three separate regions based on the primary colors. This aids in narrowing down the region where the tumor cells may be present. Iterations are done, specifically on each cluster to conclude the presence or absence of abnormal cell growth. K-means clustering algorithm (Hartigan

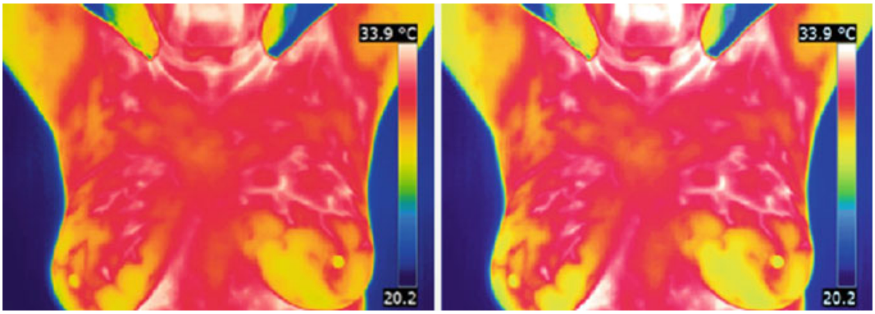


Fig. 8 Initial thermal image and contrast-enhanced image (*Source* Results from this study)

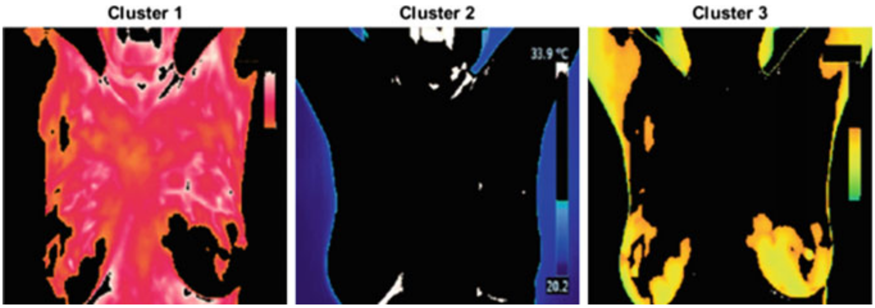


Fig. 9 Three-way clustering of thermogram using K-means algorithm based on RGB model (*Source* Results from this study)

and Wang 1979) (Lloyd 1957) (MacQueen 1967) is an investigatory algorithm in data analysis which subtracts the required area from the background and divides the image into layers, separating the hotter region from the background image [7]. Hot spot detection on thermal images becomes easier using the algorithm and it locates K-means value throughout a data highlighting its representation. It is an iterative algorithm in which K-means values are spread throughout the set in accordance with its Euclidean mean value (Fig. 9).

4.3 Feature Extraction

A thermal image has copious features that are extracted and analyzed to come to a definitive conclusion about the nature of the tumor. Features that include mean, variance, standard deviation, energy density, entropy, contrast, skewness, kurtosis, smoothness, and many others are extracted using curvelet or ridgelet transform. This transform has surpassed its predecessor in segregating the features more thoroughly and not missing any abnormal value, thereby making it a useful tool for collecting

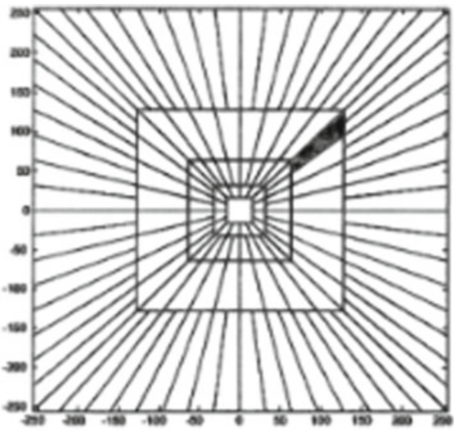


Fig. 10 Depiction of anisotropic decomposition using curvelet transforms (*Source* International Journal of Biomedical Imaging ID 136034, Fig. 8)

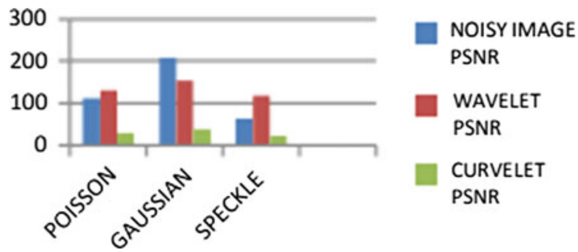


Fig. 11 Comparison of curvelet and wavelet de-noising capabilities (*Source* International Journal of Computational Engineering Research, vol. 04)

the feature database. Curvelet transform is a multiscale geometric wavelet transform, and it is more efficient in representing edges, curves, and singularities than traditional wavelet. Curvelet combines the linearity of the wavelet transform along with its own layering procedure to achieve the optimal rate of convergence by simple thresholding. Multiple decompositions slice the image into innumerable segments, thereby capturing point discontinuities in the linear structures. Curvelets in addition to a variable width have a variable length and a variable anisotropy giving it the ability to detect and measure features that are harder to notice in plain sight [6]. The image is decomposed into slices in each of the layers and the discontinuities are calculated comparing the linearity and nonlinearity of the neighboring values (Figs. 10 and 11).

The only drawback in a thermal image is that its noise level is high due to environmental factors. Also, obtaining a high-contrast image with one screening is difficult. The use of curvelet transform helps in de-noising the thermal image and the latter is resolved by the contrast enhancement. Moreover, by looking at the segmentation and layering of the curvelet transform, it gives better results in feature extraction as well as noise filtering when compared to wavelet transform [8].

The features that are finally extracted and are instrumental in the classification process are as follows:

- THERMAL DENSITY: Thermal density is the measure of difference between the textural densities in the thermal image and the complementary function.

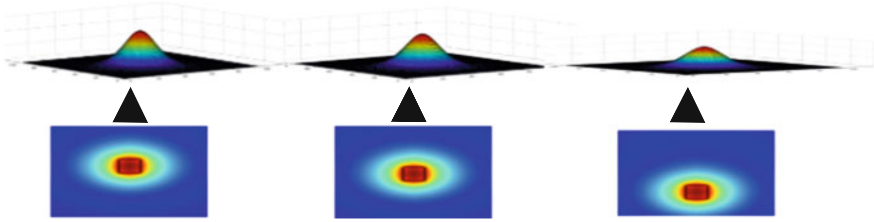


Fig. 12 Grouping of feature descriptors into high, medium, and low probability (*Source* Results from this study)

BW1 = binary thermal image,
BW2 = binary image area for the complementary function image,
Total1 = bwarea(BW1),
Total2 = bwarea(BW2), and
Thermal density = total2-total1.

- ENERGY DENSITY: The energy is expected to be high if there is a high probability of repeated pixel pairs, where $P(i, j)$ are the pixel values at the (i, j) coordinates of the image.

$$\text{Energy} = \sum_i \sum_j \{P(i, j)\}^2$$

- ENTROPY: Entropy measures the randomness of a gray-level distribution.

$$\text{Energy} = \sum_i \sum_j P(i, j) \log P(i, j)$$

- Homogeneity,
- Skewness,
- Kurtosis,
- Mean and variance, and
- Contrast.

Similar descriptors are grouped together into arrays that form the testing database which is further compared with a training database in accordance with the support vector (Fig. 12).

5 Classifications for Thermogram and 3D Mammogram

Classification techniques usually consist of taking a biopsy and analyzing it for abnormalities that suggest the presence of cancer or benign cells. The traditional method for classification is taking a biopsy of the infected tissue and analyzing it for the nature, cause, and effects of the tumor. This procedure is highly invasive and takes prolonged time to provide a conclusive answer. Therefore, the fastest kernel classification machine learning method called fast/core support vector machine is used to train the medical database with the acquired features extracted [9]. These existing values are compared and correlated with the features and values extracted from the curvelet transform and wavelet transform for thermogram and mammogram, respectively. An SVM train is added to the existing image and the computer calculates the probability of malignant cells present. An SVM vector is calculated and the probabilities are placed over the vector in regions of high, low, and reference [10]. Values that are proximal to the vector, either high or low, are considered as manifestations of malignancy.

The various classification outputs that can be derived are as follows [11]:

- True positive: This result confirms the cells as malignant and the SVM computes the affected area and accuracy of calculation by multiple iterations.
- True negative: This result confirms the tumor cells as benign and prompts the computer to analyze a different cluster in the odd chance of the cells being present in other layers and regions of tissue.
- False positive: This is an unacceptable result that points out the presence of malignant cells when in reality, they are benign.
- False negative: The SVM algorithm fails to detect the presence of cancer cells because their manifestation is not very prominent and the features do not match with the exact threshold values of the training database.

The FSVM is used to make sure that the classification provides satisfactory results with least number of iterations and high accuracy.

5.1 Thermogram

After the classification result finalizes the presence of cancer, identification of the exact location of the cells is carried out. This is performed by minimum enclosing ball (MEB) algorithm. The malignant cells may not only be present over the surface tissue layer but it may have also penetrated into the interior walls. MEB is used to detect a boundary of infected cells and sequester them in the shape of spheres [12]. This spherical area is decomposed and further layers are inspected for abnormality and a graphical output is presented depicting the exact cell locations in the entire thermogram (Fig. 13).

5.2 Mammogram

Various classifiers are used on mammograms like Parzen's classifier and linear discriminant classifier (LDC). It has been found out that best results are obtained when

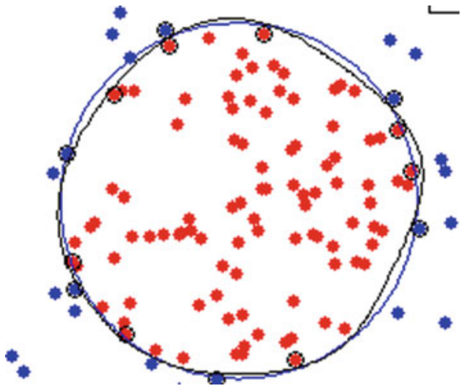


Fig. 13 MEB boundary detection. Red dots represent malignant cells, blue dots are benign cells (*Source* Result from this study)

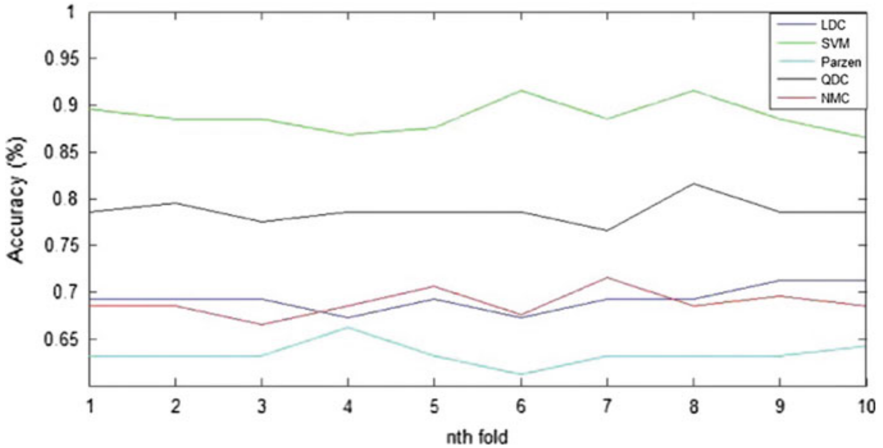


Fig. 14 Graph of accuracies for various classifiers when SWT is used for feature extraction (*Source* Technol Cancer Res Treat, p. 510)

SWT and SVM are combined [5]. After classification as benign or malignant, further boundary detection processes are not yet developed for mammograms (Fig. 14).

6 Results

After the acquisition, processing, segmentation, feature extraction, and classification techniques have been performed on both types of images, thermogram and 3D mammogram, a conclusive result with accuracy has been coined. For the mammogram, when SWT features are combined with SVM, a maximum accuracy of 88.80% is achieved while DWT features in combination with SVM yielded a maximum accuracy of 81.73%. Similarly, for an infrared image, curvelet transform features combined with SVM classification yielded an accuracy of 95.16%, sensitivity of 96.77%, and specificity of 98.38%.

7 Conclusions

With the cumulative results from both images, we can conclude that thermogram combined with SVM classifier provides a better accuracy in distinguishing the nature of the tumor, consequentially paving way for earlier detection as well as diagnosis and treatment. Though mammograms are the most sought after and reliable techniques for detection, the follow up processes do not give satisfactory results. The ultimate goal is to provide a platform for women of all ages and walks of life to be able to detect the cancer at an early stage and get the necessary help they need as soon and as effectively as possible. This procedure not only revolutionizes the field of medicine but also bridges the gap between image processing, medicine, data sciences, and machine learning.

References

1. Gonzalez RC, Woods RE. Digital image processing, 3rd edn. Pearson Education
2. Akinyemiju TF (2013) Risk of asynchronous contralateral breast cancer: multiple approaches for a complex issue. Gland Surg 2(2):110–113
3. Cowley G (2017) Mammography vs. thermography: comparing the benefits. Medical News Today
4. Geoge KMJ, Dhas DAS (2017) Preprocessing filters for mammogram images: a review. In: Emerging devices and smart systems (ICEDSS). ISBN: 978-1-5090-5555-5. IEEE Xplore 19 Oct 2017
5. Ganesan K, Acharya UR, Chua CK, Min LC, Abraham TK (2014) Automated Diagnosis of Mammogram images of breast cancer using discrete wavelet transform and spherical wavelet transform features. Technol Cancer Res Treat 13(6):605–615
6. Mahmoudzadeh E, Zekri M, Montazeri MA, Sadri S, Dabbaggh ST (2016) Directional SUSAN image boundary detection of breast thermogram. IET Image Process 552–560
7. Shmmala FA, Ashour W (2013) Color based image segmentation using different versions of K-means in two spaces. Global Adv Res J Eng Technol Innov 1(9):030–041. (ISSN: 2315-5124)
8. Shukla M, Changlani S, A Comparative study of wavelet and curvelet transform for image
9. Catanzaro BC, Sundaram N, Keutzer K (2008) Fast support vector machine training and classification on graphic processors. In: UCB/EECS
10. Phan J, Moffitt R, Dale J, Petros J, Young A, Wang M (2005) Improvement of SVM algorithm for microarray analysis using intelligent parameter selection. In: 2005 IEEE-EMBS 2005 27th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 4838–4841
11. Shrivastava P, Singh P, Shrivastava G (2014) Image classification using SOM and SVM feature extraction, (IJCSIT) 5(1):264–271
12. Cervantes J, Li X, Yu W, Li K (2008) Support vector machine classification of large data sets via minimum enclosing ball clustering. Neurocomputing 71(4–6):611–619

IEFA—A Fuzzy Framework for Image Enrichment

Ankita Sheoran and Harkiran Kaur

Abstract In this work, an Image Enhancement Fuzzy Algorithm (IEFA), a technique for image enhancement has been proposed and developed. IEFA formulates the mapping from a given input to an output using fuzzy logic. IEFA improves the contrast of low-contrast images. The technique begins the process of image enrichment by modifying membership functions and designing fuzzy if–then rules that exist as a sophisticated bridge between human knowledge on one side and the numerical framework of the computers on the other side. The algorithm converts image properties into fuzzy data and further fuzzy data into crisp output through defuzzification. Further, to evaluate the performance of the proposed technique, the developed technique has been compared with "Histogram Equalization (HE) and Contrast Limited Adaptive Histogram Equalization (CLAHE)." It has been observed that PSNR and CII of the proposed algorithm (using a test image) are 25.56 and 1.13, respectively. These metrics are 0.078 and 6.603% more effective than the metrics of existing algorithms.

1 Introduction

Present-day applications require various kinds of images and pictures as sources of information for interpretation and analysis. On applying conversion of an image from one form to another form, such as digitized, scanned, transmitted, stored, and so on, some of the degradation occurs at the output state. Hence, the image has to undertake the process of image enrichment which consists of collection of techniques that seek to improve the visual appearance of an image.

Image enhancement techniques have wide number of applications in the field of medical science imaging, art studies, forensic science, and atmospheric science.

A. Sheoran · H. Kaur (✉)
Computer Science and Engineering Department, Thapar University, Patiala 147001, Punjab, India
e-mail: harkiran.kaur@thapar.edu

A. Sheoran
e-mail: ankita.shrn@gmail.com

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_122

Image enhancement improves the quality of the image so as to prevail over the limitations of the human visual system. The incentive of image enhancement techniques includes higher visual quality, extracting the hidden details in the image, increasing the contrast of low-contrast image, enhancing image features for further processing, and many more. In image enhancement, the low-quality image is transformed into high quality with the intent of improving the look of the image [1]. Instead of increasing the inherent information of the data, enhancement upsurges the range of features.

1.1 Image Enhancement (IE)

In IE process, the visual look of image is perked up or is converted in a form that can be easily understood by human eyes or machines. Noisy image data is very difficult to enhance, which is necessary for many research and application areas. There are three main categories in image enhancement technique which are given as follows:

i. **Spatial domain methods**

"Spatial domain method" functions directly on pixels. In spatial domain method, according to rules, the pixel values are adapted that are dependent on the original pixel value (local or point processes). On the other hand, in many ways, the pixel values can be merged with or compared to other pixels within instant neighborhood. Consider the input image $r(j, k)$ and processed image $s(j, k)$, then the transformation $s(j, k) = Q[r(j, k)]$, where Q is an operator on r defined over some region of (j, k). The operator Q is applied at every position (j, k) to yield output s at that position. The process uses pixels in the area of image spanned by vicinity [2]. The two most popular conventional methods in spatial domain are histogram: specifications and equalization, and adaptive neighborhood histogram equalization [3].

ii. **Frequency domain methods (DFT)**

In frequency domain method, Fourier transformation of image is used. Transformation of image into two dimensions even with fast transformation is a very time-consuming task, thus making it less appropriate for real-time processing [3]. Sharp conversions and edges in image give extensively to high-frequency content of Fourier transformation [2]. The overall appearance of the image over smooth areas is due to low-frequency content in Fourier transformation. It is easy to study the idea of filtering in frequency domain and so improvement of image $r(j, k)$ can be completed in the frequency domain based on DFT [2]. This is chiefly beneficial in convolution if the spatial extent of the point spread sequence $h(j, k)$ is larger than the convolution theory.

$$s(j, k) = h(j, k) * r(j, k) \tag{1}$$

where $s(j, k)$ is enhanced image and $r(j, k)$ is input image.

iii. **Fuzzy domain method**

Fuzzy set theory is suitable for handling diverse uncertainties in image processing and computer vision applications. Fuzzy approaches comprise different sets that help in recognizing, characterizing, and processing of image which as a whole describes the fuzzy theory. Fuzzy logic has three processes: image fuzzification of crisp value to fuzzy values, membership function modification, and defuzzification to get back the crisp values. Fuzzy image enhancement involves mapping of gray level into membership function with the intent of generating an image of greater contrast than the input image. This is achieved by assigning a larger influence to the pixel intensity that is nearer to the average pixel intensity and to those that are beyond the average intensity of the image.

1.2 Fuzzy Logic

It is a multivalued logic in which variables can have truth values in the range 0–1 and can be any real number, whereas the truth values of variables in the classical Boolean logic can only be the "crisp" values "0 or 1" that is "completely true" or "completely false." Fuzzy logic is engaged to deal with the impression of partial truth, in which the truth value may vary between "completely true" and "completely false" in which membership function manages the degree. Process:

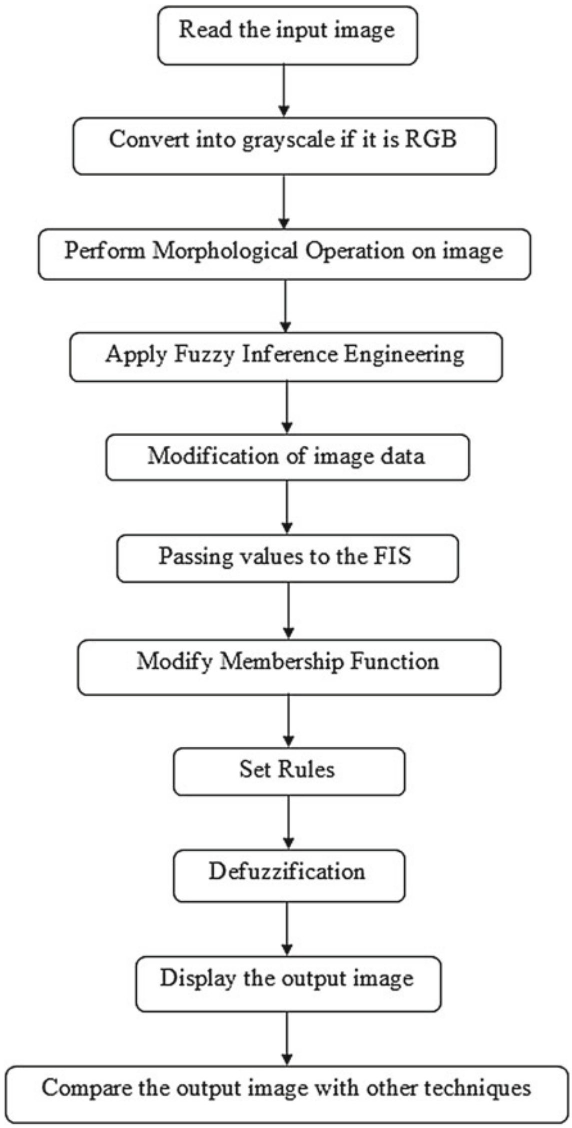
(a) Fuzzifying input values that are pixel data into fuzzy membership functions.
(b) Implement all applicable rules from the Rule Base to calculate the fuzzy output functions.
(c) Defuzzify the output membership functions for "crisp" output values.

Fuzzy logic systems can perform well even with imprecise and incomplete input data or even if data is not reliable, because in fuzzy logic the output is an agreement of the inputs and all the defined rules in the Rule Base [4]. By optionally adding weights to each rule, the degree up to which it affects the output can be controlled. These rule weights can also be based on outputs of other rules. The weights can be changed or can be kept static. The criterion for deciding the weights is consistency or reliability and how important the rule is.

2 Design and Implementation of Framework

In this work, the "Mamdani Fuzzy Inference System" is chosen over "Sugeno Fuzzy Inference System" because the system has to take care of three diverse ranges of pixel strength, namely, bright, dark, and gray. In Sugeno FIS, it is beyond the bounds of possibility to unite rules designed for the system as single spike is used for every designed rule.

2.1 Workflow of the System



2.2 Description of Image Enhancement Fuzzy Algorithm (IEFA)

Step 1: Morphological processing which includes the following:

i. Convert the image into grayscale if it is RGB.
ii. Convert the image into double format.
iii. Evaluate size of the image.
iv. Calculate minimum, maximum, and mean gray level intensity of the image.

Step 2: Transforming the image data into fuzzy domain data that is performing fuzzification.

The FIS toolbox of MATLAB comprises FIS editor, rule editor, rule viewer, and membership function editor. The FIS editor shows general information of FIS. It displays the input and output variables. Here, the input variable is "pel_in" and output variable is "pel_out." The "ALGORITHM I" converts the image pixel data into fuzzy data. The "pel_in" takes the output of ALGORITHM I, that is, fuzzified pixel value.

Logical Structure of Algorithm I

```
FOR grayvalue = each pixel of the image
IF (grayvalue is between 0 AND minimum)
fuzzy=0;
  ELSE IF (grayvalue is between minimum AND mean)
    fuzzy = (1/(mean - minimum)) * minimum + (1/(mean - minimum)) * grayvalue;
  ELSE IF (grayvalue between mean and maximum)
    fuzzy = (1/(maximum - mean)) * mean + (1/(maximum - mean)) * grayvalue;
  ELSE (grayvalue between maximum AND 255)
    Fuzzy = 1;
  END
END
```

The low-contrast input image is transformed to membership plane with membership function where its values lie in the range 0–1. The membership function adopts any value in the interval 0–1. Three triangular membership functions are taken for both input and output at three intensity levels of the image, respectively. For each membership function, range is adjusted in membership function editor toolbox of FIS.

Step 3: Membership modification.

Using Fuzzy Inference System (FIS) toolbox, the membership values are modified. FIS consists of inbuilt membership functions for both input and output variables. New membership values are defined for the pixels with respect to their input pixel intensity, and based on variation in their intensity level the membership values are assigned. The membership values are modified for the pixels considering their gray level intensity values, and hence membership values are allotted to the pixels on the basis of variation in their intensity level.

Logical Structure of Algorithm II

```
FOR pixel = every pixel value
   IF (pixel_value is between 0 AND 0.5)
new_pixel = 5 * (pixel ^ 3 );
   ELSE
new_pixel = 1 − 5*((1 − pixel) ^ 3);
   END
```

Step 4: Set rules.

The rules are constructed in the graphical rule editor interface of fuzzy inference system tools. Fuzzy rules are defined for the modified membership values with respect to the pixel intensity. The ranges of both output and input membership functions have been considered in defining the fuzzy rules. The rules are added by taking into account the parameters of the selected triangular membership function for the system. The designed IEFA has been designed and implemented on the basis of the following fuzzy rules:

i. If pel_in is Dark then pel_out is More_dark.
ii. If pel_in is Gray then pel_out is More_gray.
iii. If pel_in is Bright then pel_out is More_bright.

As a whole, the input of the if–then rule is the present value of the input variable (pel_in) and the output is the whole fuzzy set (More_dark, More_gray, or More_bright).

Step 5: Convert fuzzy data into grayscale enhanced data that is performing defuzzification.

Defuzzification is the final step in Mamdani FIS. For defuzzification process, combine output fuzzy collection and the final output is a single crisp number. The image is converted to the pixel plane from fuzzy plane. The output of Algorithm III is crisp number that enhanced pixel values.

Logical Structure of Algorithm III

Set maximum intensity and minimum intensity for enhanced image as:

```
maxI = 255;
minI = 0;
FOR new_pixel = each pixel value
   IF (pixel <= minimum)
pixel_enhanced = 0;
   ELSE IF (pixel is between minimum AND maximum)
pixel_enhanced = ((maxI - minI) * result(x,y) + minI);
   ELSE
pixel_enhanced = 255;
   END
END
```

3 Experimental Setup

To assess the proposed technique, different low-contrast images have been taken. The following image enhancement techniques have been applied to the input images. The techniques are given as follows:

3.1 Histogram Equalization Technique

Figure 1 shows the original low-contrast image and the histogram representation of it.

Figure 2 shows histogram equalized image and its histogram. Histogram equalization technique evens out the plot so that it covers the complete range of brightness with almost similar amount of pixels for every brightness range. Also, complete range of brightness value is covered and the graph is balanced.

3.2 Contrast Limited Adaptive Histogram Equalization (CLAHE)

"CLAHE" is different from the ordinary HE in respect that the adaptive technique equalizes numerous histograms, each of which corresponds to a different segment

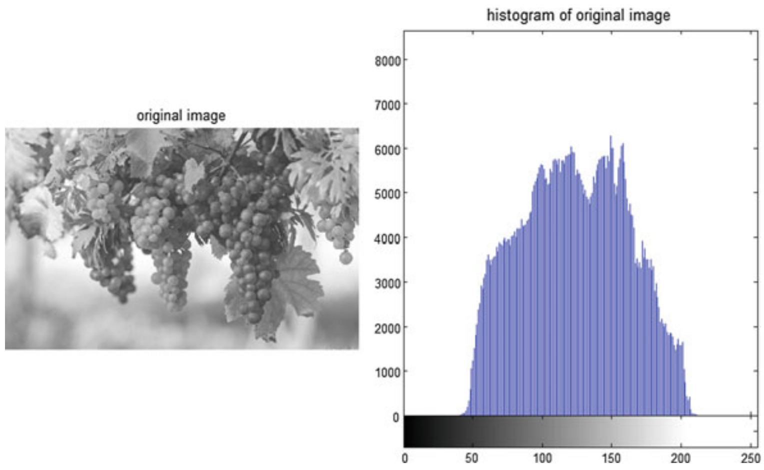


Fig. 1 Original low-contrast Grapes1.jpg and its histogram plot

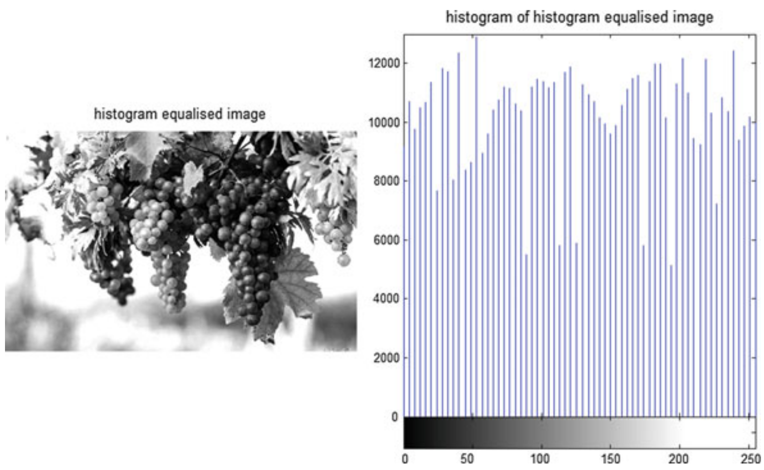


Fig. 2 Histogram equalized Grapes1.jpg and its histogram plot

of the image and redistribute the lightness values of the image using them. Figure 3 shows the CLAHE improved image.

3.3 Image Enhancement Fuzzy Algorithm (IEFA)

The proposed work aims at improving the contrast of low-contrast image by reducing the noise, without overenhancing the contrast. The method is based on intensity-level mapping onto fuzzy plane with the membership function transformation that

Fig. 3 Applying CLAHE on Grapes1.jpg

Fig. 4 Applying IEFA on Beach1.jpg

Fig. 5 Applying IEFA on Rock1.png

increases the contrast of the original low-contrast image. Figures 4, 5, 6, and 7 display the visual outcomes attained after applying IE. The radiology images in Figs. 6 and 7 were taken from radiology department of a local hospital.

4 Performance Evaluation

After applying the abovementioned image enhancement techniques on low-contrast images, the results have been obtained. The performance analysis of various tech-

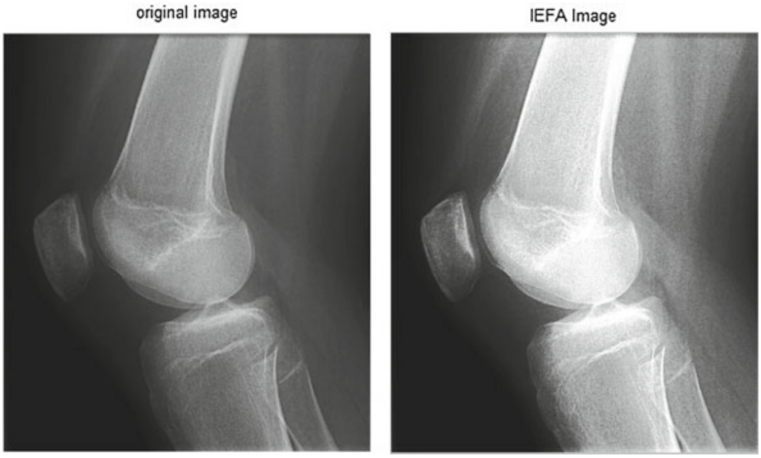


Fig. 6 Applying IEFA on Medical1.jpg

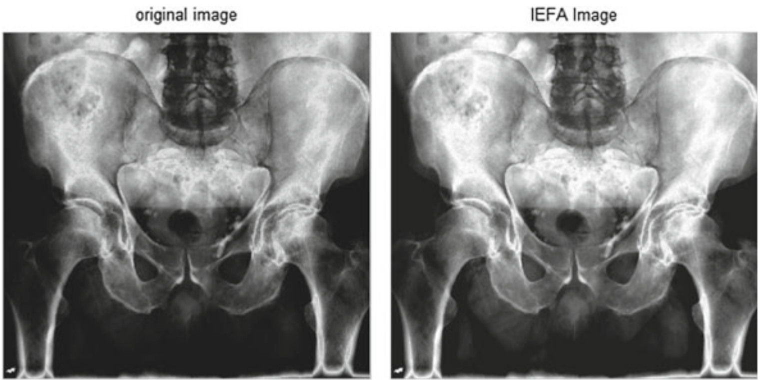


Fig. 7 Applying IEFA on Medical2.jpg

niques has been done using different parameters that is “Peak-signal-To-Noise Ratio (PSNR)” and “Contrast Improvement Index (CII).”

PSNR

“PSNR is the division of the maximum possible power of a signal to the power of corrupting noise affecting image quality.” For reconstruction to be of higher quality, the value of PSNR should be high. Table 1 shows comparison of PSNR of enhancement techniques.

CII

Contrast improvement index is the division of C_{enhanced} to C_{original}, whereC_{enhanced} is average contrast of new image and C_{original} is the average contrast of input image. A higher value of CII is always favored. Figure 8 displays a bar representation of CII

Table 1 Comparison of PSNR of various enhancement techniques

Image name	PSNR of the original image	PSNR of histogram equalized image	PSNR of CLAHE image	PSNR of IEFA
Grapes1.jpg	25.25	25.56	25.34	25.56
Grapes2.png	25.26	25.57	25.41	25.55
Beach1.jpg	25.27	25.54	25.34	25.56
Rock1.png	25.25	25.55	25.41	25.55
Field1.png	25.27	25.54	25.35	25.58
Field2.jpg	25.25	25.55	25.33	25.55
Field3.png	25.27	25.54	25.37	25.57
Field4.jpg	25.25	25.54	25.39	25.54
Medical1.jpg	25.30	25.56	25.28	25.43
Medical2.jpg	25.34	25.52	25.45	25.41

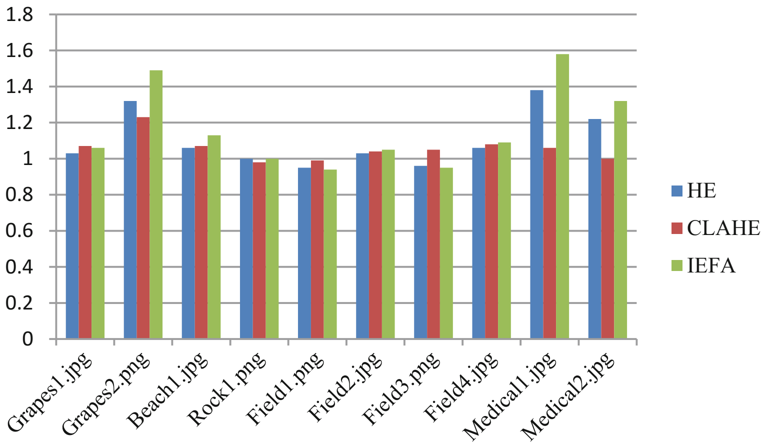


Fig. 8 CII evaluation of HE, CLAHE, and IEFA

achieved by applying the techniques on low-contrast images. From this figure, it is inferred that proposed IEFA yields the highest value of CII (Table 2).

5 Conclusion

The foremost focus of IEFA is on image enhancement of low-contrast images using fuzzy technique. For implementation, three membership functions are defined for each input and output, and rules are designed in FIS. Experiment has been performed on low-contrast images and the outcome of the proposed IEFA is compared with the

Table 2 CII analysis

Image name	HE	CLAHE	IEFA
Grapes1.jpg	1.03	1.07	1.06
Grapes2.png	1.32	1.23	1.49
Beach1.jpg	1.06	1.07	1.13
Rock1.png	1.00	0.98	1.00
Field1.png	0.95	0.99	0.94
Field2.jpg	1.03	1.04	1.05
Field3.png	0.96	1.05	0.95
Field4.jpg	1.06	1.08	1.09
Medical1.jpg	1.38	1.06	1.58
Medical2.jpg	1.22	1.00	1.32

outcome of HE and CLAHE. To contrast these algorithms, "Contrast Improvement Index (CII)", and "Peak-Signal-To-Noise Ratio (PSNR)" have been used as the performance metrics. From CII analysis, it is concluded that "Histogram Equalization (HE)" and "Contrast Limited Adaptive Histogram Equalization (CLAHE)" yield less values of CII. This is because the focus of HE is only on the global contrast of image. This also leads to the loss of the local details. In CLAHE, because of nonuniform lightning, ambiguity is introduced in image which appears as imprecise boundaries during digitization. IEFA results in better CII values as compared to CLAHE and HE by solving the problem of vagueness and imprecise boundaries through fuzzy sets and linguistic variables.

5.1 Future Scope

In future, the work can be extended to colored images that offer more real-life implementations. The algorithm can be modified to give better results for medical images. The improvised outcome for image enrichment is also used in real-time augmentation of neuroevolution.

References

1. Yaman O, Karakose M (2016) Development of image processing based methods using augmented reality in higher education. In: 15th International conference on information technology based higher education and training (ITHET). IEEE
2. Patil M (n.d) Design of novel fuzzy based method for contrast. Int J Electr Electron Data Commun. ISSN 2320-2084
3. Kaur T, Sidhu RK (2016) Optimized adaptive fuzzy based image enhancement techniques. Int J Sig Process Image Process Pattern Recogn 9.1:11–20
4. Tizhoosh HR (2000) Fuzzy image enhancement: an overview. Fuzzy techniques in image processing. Physica-Verlag HD, pp 137–171

Hidden Markov Random Field and Gaussian Mixture Model Based Hidden Markov Random Field for Contour Labelling of Exudates in Diabetic Retinopathy—A Comparative Study

E. Revathi Achan and T. R. Swapna

Abstract Diabetic Retinopathy (DR) is one of the important causes of blindness in diabetic patients. Diabetes that affects the retina is called diabetic retinopathy. Diabetic retinopathy occurs due to the damage of blood vessels in retina and increase in the level of glucose. Different pathologies are normally seen in DR such as microaneurysms, hard exudates, soft exudates, cotton wool spots and haemorrhages. We have done a comparative study of Hidden Markov Random Field (HMRF) and Gaussian Mixture Model (GMM) based HMRF for automatic segmentation of exudates and the performance analysis of both methods. The preprocessing consists of candidate extraction step using greyscale morphological operation of closing and initial labelling of exudates using K-means clustering followed by contour detection. In contour detection, we have analysed two approaches, one is GMM-based HMRF and the other is HMRF. DIARETDB1 is the dataset used.

1 Introduction

Diabetes is a metabolic disease which occurs due to a high level of sugar. The high level of sugar produces symptoms like polymia, polydipsia and polyphagia. DR is the diabetic eye disease which occurs due to damage to the retina and may cause blindness. The treatment and monitoring of retina are important to diagnose and monitor the progress of the disease as there are no warning signs for this disease. There are two types of diabetic retinopathy such as nonproliferative diabetic retinopathy and proliferative diabetic retinopathy [1]. At the stage of nonproliferative diabetic

E. Revathi Achan · T. R. Swapna (✉)
Department of Computer Science and Engineering,
Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Coimbatore, India
e-mail: tr_swapna@cb.amrita.edu

E. Revathi Achan
e-mail: cb.en.p2cvi16006@cb.students.amrita.edu

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_123

retinopathy, the leakage of blood vessels occur. In proliferative diabetic retinopathy, abnormal blood vessels are formed at the back of the eye and can burst or bleed and can cause vision loss. The pathological signs of DR are microaneurysms, haemorrhages, exudates and cotton wool spots [2]. There are different modalities used to identify diabetic retinopathy such as colour fundus photography, fundus fluorescein angiography and optical coherence photography. Fundus photography is used to examine the retina by biomicroscope with magnifying lens. Fundus angiogram is used to detect the leakage in blood vessels. Optical coherence tomography is a light-sensitive tissue to sense at the back of the eye which uses light waves to detect the abnormality in retinal images. It gives the depth information of the image. The use of light gives more axial resolution than any other imaging technique.

2 Literature Survey

Authors	Methodologies	Outcomes	Remarks
Hussain F. Jaafar et al. [3]	1. Coarse and fine segmentation 2. Split and merge technique	99.4% of accuracy	Due to distinctive performance measures, the proposed method may be successfully applied to images of variable quality
Choundhury S et al. [4]	1. Fuzzy c means based feature extraction 2. Preprocessing step includes smoothening and green channel of image	Using SVM the exudates of 97.6% accuracy is obtained	Detected only blood vessel density
Harini R., Sheela N. [5]	1. Preprocessing step techniques such as image resizing, contrast enhancement, grey and green colour extraction 2. Fuzzy c means and morphological operation	Accuracy of 99.67%, sensitivity of 100%, specificity of 95.83%	Features collected have considered for classification are area, microaneurysms, haemorrhages
Choundhury S et al. [4]	1. Machine learning techniques 2. Classifiers such as Gaussian mixture model, KNN, GMM and support vector machines	53 images contain exudates	More time complexity and future work can be directed towards detection of neovascularization

(continued)

(continued)

Authors	Methodologies	Outcomes	Remarks
R. Priya, P. Aruna [6]	1. SVM and neural network methods 2. Area, perimeter, radius of features are captured	The performance measures of SVM have an accuracy of 97.608% and probabilistic neural network has an accuracy of 89.60%	Mainly, SVM outperforms the other model
Ramon Pires [7]	1. Features are extracted and given to the input of SVM and bossa nova technique	Detection performance of 96.4% for hard exudates and 93.5% for red lesions	Bossa nova is used for mid-level feature characterization and image representation for different lesion detectors
Raju Maher et al. [8]	1. Preprocessing step consists of image contrast enhancement such as histogram equalization, morphological operators followed by binarization	That 96.9% of sensitivity, 96.1% specificity and 97.38% of accuracy	Automatic identification of image processing techniques for abnormalities in retinal images
Kusakunniran W et al. [9]	1. Four features are extracted such as colour, contrast, focus and illumination 2. Grabcut algorithm	Image segmentation based on the iterative selection and grabcut algorithm	Images are qualified on the basis of quality assessment and segmentation of diabetic retinopathy images
B. V Shilpa, T. N Nagabhushan [10]	1. Novel-based ensemble approach 2. Morphological operations compared with logical operations	89.13% is the positive predicted value 100% is the negative predicted value	Use of canny edge detection algorithm has further improved the accuracy of boundary detection, along with the removal of false connected components
Arslan Ahmed et al. [2]	1. Green channel, histogram equalization, Morphological smoothening are the preprocessing techniques 2. KNN and SVM classifiers are used	The proposed method has sensitivity of 97.39%, specificity of 98.02%, accuracy of 97.56% and AUC of 0.97%	These methods are using hybrid classifier
Mahendra Gandhi, Dr. R. Dhanasekharan [11]	1. Converting the image into greyscale and edge detection algorithm 2. Features are extracted using GLCM and given to the input of SVM	SVM classifier is used to detect exudates	This paper not only confirms the disease but also tends to measure the severity of the disease
Ganesh S, Dr. A.M. Basha [12]	1. Krisch's algorithm is used and classified	Blood vessel extraction from retinal images was done	All the features are extracted and analysed but the correct verification was not done

(continued)

(continued)

Authors	Methodologies	Outcomes	Remarks
Sindhura et al. [13]	Preprocessing step such as closing operation, RGB to HSI, adaptive histogram equalization	Percentage of exudates detected is 40.16%	Nilblack's thresholding is used and superimposed the image onto the original image for showing exudates
Shevta, Gurmeen Kaur [14]	Support vector machine and multilinear discriminant analysis	Detection of exudates was done up to 50%	The review for detection of DR using fundus images and approaches SVM and MDA
Anup. V. Deshmukh [15]	Enhanced using brightness transform function	Achieves 99% of accuracy	The local mean and entropy-based region growing technique is applied to classify exudate and non-exudate pixels
Amrita Roy Chowdury, Sreeparna Banerjee [16]	Contrast enhancement, optical disc is detected and segmented using fuzzy c means	The automatic detection can help the doctors for accurate detection of cotton wool spots and also longitudinal study of retinal image	All the spots are not detected properly
Rakshitha T. R. et al. [17]	1. Wavelet transform method 2. Techniques are examined using PSNR ratio	Calculating the peak-to-signal ratio from datasets	Best results of detection of edges compared to other techniques
Yitian Zhao et al. [18]	New saliency-based technique	Detection of leakage in fluorescein angiography	The experimental results show that it outperforms one of the latest competitors and performs leakage detection
Meenu Vijayan et al. [19]	Life assessment between patients based on questionnaire method	Evaluated the cost for different treatment modalities	Study on quality of life assessment in diabetic retinopathy among patients
Parvathy Ram, Swapna T. R. [1]	A method for profiling of hard exudates and microaneurysms	Quantification algorithms are used	Not detected automatically and used the crop operation to detect the exudates

3 Methodology

The methodology of the approach is as shown in Fig. 1. The standard database DIARETDB1 is used as an input image to detect the exudates. It consists of 89 colour fundus images. The database consists of original images and ground truth images of soft and hard exudates. The image is captured at 50° field of view using fundus camera. This dataset is defined as calibration level 1 fundus images. In the candidate extraction step, the exudate regions are extracted using greyscale morphological closing with size 3×3 structuring element. Closing operator is used to enlarge the boundaries of foreground and also preserves the background region. In contour detection step, we have used both HMRF- and GMM-based HMRF. HMRF comprises two algorithms, expectation maximization algorithm and MAP estimate.

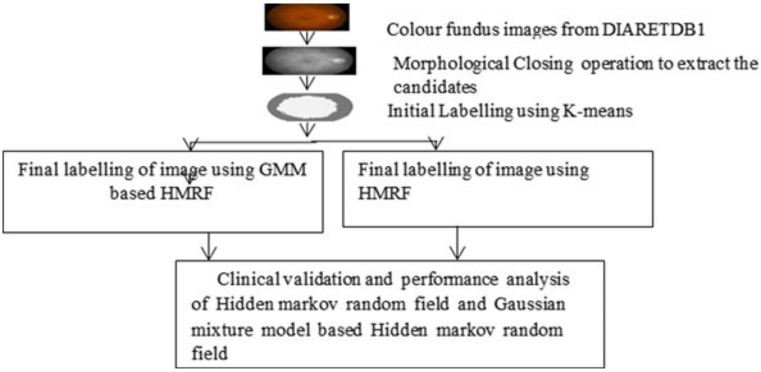


Fig. 1 Methodology

Using expectation maximization, the intensity distribution of each region is segmented using Gaussian distributions with parameters such as mean and covariance denoted by $\theta_{x_i} = (\mu_{x_i}, \sigma_{x_i})$.

3.1 Algorithm for Detection of Exudates

(a) Read an input image,
(b) Candidate extraction by grayscale morphological closing,
(c) Initialize the foreground and background pixels by HMM- and GMM-based HMM using MAP estimation and expectation maximization algorithm to find the exact boundaries.

3.2 Hidden Markov Random Field-Expectation Maximization Algorithm for Contour Detection

For the given image $\mathbf{Y} = (y_i \ldots y_n)$ and for each y_i is the intensity of a pixel, configuration of labels is defined in $\mathbf{X} = (x_1 \ldots x_n)$, where $x_i \, \varepsilon \, L$. L is a set of labels. Expectation maximization algorithm is used to estimate the parameters $\boldsymbol{\theta} = \{\theta_l | l \, \varepsilon \, L\}$.

(a) Initial parameter set $\boldsymbol{\theta}^{(0)}$ and in Expectation (E) step at tth iteration conditional expectation is

$$Q(\boldsymbol{\theta}|\boldsymbol{\theta}^t) = \sum_{x \varepsilon \chi} P\big(X|Y, \boldsymbol{\theta}^{(t)} \ln P(X, Y|\boldsymbol{\theta})\big), \tag{1}$$

where chi represents the configuration of labels.

(b) In M-step: Now maximize $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)})$ to obtain the next estimate

$$\boldsymbol{\theta}^{(t+1)} = \text{argmax } Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)}) \tag{2}$$

Then, let $\boldsymbol{\theta}^{(t)} \rightarrow \boldsymbol{\theta}^{(t+1)}$ and repeat from the E-step.

(c) $G(z; \theta_l)$ represents the Gaussian distribution function with parameters

$$\theta_l = (\mu_l, \sigma_l)$$
$$G(z; \theta_l) = \frac{1}{\sqrt{2\pi\sigma_l^2}} \exp(-(z - \mu_l)^2/2\sigma_l^2). \tag{3}$$

(d) Estimate the labels using current parameter set by MAP algorithm.

(e) Calculate the posterior distribution of all $l \varepsilon L$ in pixels y_i.

(g) Update the parameters of $(\mu_{l,} \sigma_l)$.

3.3 Maximum A Posteriori Algorithm of HMRF

(a) Already we have an initial estimate x^0,which we observed from EM algorithm.

(b) Provided x^k for all $1 <= i <= N$,

$$U(\boldsymbol{Y}|\boldsymbol{X}, \boldsymbol{\theta}) = \sum [(y_i - \mu_{x_i})^2/2\sigma_{x_i}^2 + \ln \sigma_{x_i}] \tag{4}$$

is the likelihood energy Eq. (4).

(c) Clique potential is defined on pairs of neighbouring pixels

$$v_c(x_i, x_j) = 1/2(1 - I_{x_i}, x_j). \tag{5}$$

(d) Repeat step 2 until $U(\boldsymbol{Y}|\boldsymbol{X}, \boldsymbol{\theta}) + U(\boldsymbol{X})$ converges or a maximum k is achieved.

3.4 Gaussian Mixture Model Based Hidden Markov Random Field

(a) Initial parameter set $\boldsymbol{\theta}^{(0)}$ and expectation step at tth iteration conditional expectation is calculated as per Eq. (1) and used to determine which data belong to which Gaussian components.

(b) In maximization step of $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)})$ is calculated as per Eq. (2) and used to recompute the Gaussian model parameters. GMM with g components can be represented using $\theta_l = (\mu_{l,1}, \sigma_{l,1}, w_{l,1}) \ldots (\mu_{l,g}, \sigma_{l,g}, w_{l,g})$. When comparing the hidden Markov model of Gaussian distribution function of $\theta_l = (\mu_l, \sigma_l)$, this model has weighted probability:

$$G_{\min}(z, \theta_t) = \sum_{c=1}^{0} w_{l,c} G\big(Z; \mu_{l,c}, \sigma_{l,c}\big). \tag{6}$$

(c) Maximum a priori algorithm for labels is estimated using $x^* = \operatorname{argmin}\{U(Y|X,\theta)+U(X)\}$, where $U(X) = \sum_{c\in C} v_c(x)$ is the clique potential. And, in the image, one pixel has almost four neighbourhood pixels. Then, clique potential is defined in Eq. (4) in the HMRF model but in the Gaussian model, this equation of constant coefficient ½ is replaced by variable coefficient as β.

$$v_c(x_i, x_j) = \beta(1 - I_{x_i}, x_j). \tag{7}$$

After initial labelling, the final labelling is done using HMRF- and GMM-based HMRF. Both are using expectation maximization algorithm and maximum a posteriori algorithm. But, the difference between them are parameters discussed above.

4 Result Analysis

In Fig. 2, the ground truth images are already labelled by medical experts as shown in Fig. 2(a–c). The colour fundus image in DIARETDB1 is the input as in Fig. 2 (d–f) and then preprocessing technique using morphological closing operation for candidate extraction is applied as in Fig. 2(g–i). K-means segmentation is used for finding initial labels and k value is initialized as three, such as one for foreground and other two for background and exudate/non-exudate portions. The final labelling is done using expectation maximization, MAP algorithm and initialized the number of iteration values as 10 to get the refined segmentation and to remove non-exudate portions of image as in Fig. 2(m–o).

As shown in Fig. 3, the DIARETDB1 is the database and preprocessing is done using grayscale closing operation for candidate extraction. Then, initialized the value of $k = 3$ (k-means) for foreground, background and exudate/non-exudate portion for segmentation. This method uses $g = 3$ (Gaussian components). The k-means are used for clustering/grouping a given data by minimizing the distance between them and the centroid. The Gaussian mixture model is used to determine the probability that a point 'g' is in cluster 'k' as in Fig. 3(j–l). In the expectation maximization step, we determine which data belong to which Gaussian component and in maximization step it recomputes the GMM parameters as shown in Fig. 3(m–o).

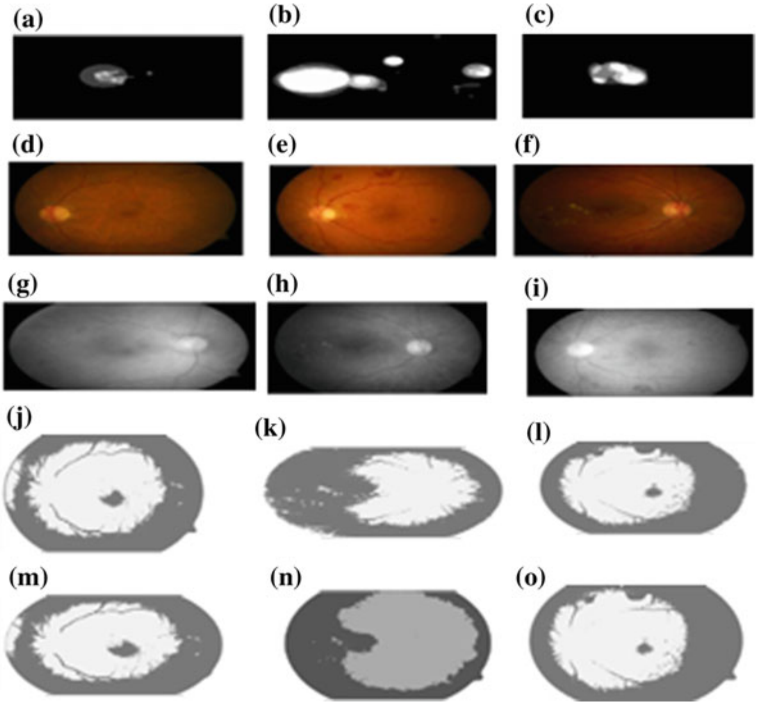


Fig. 2 **a–c** Ground truth images, **d–f** DIARETDB1, **g–i** closing operations, **j–l** initial labelling using k-means, **m–o** HMRF

As per Table 1, DIARETDB1 consists of 89 images in total with dimension 1500 × 1152. Ground truth images are provided for DIARETDB1, where hard and soft exudates are marked by various experts. The pixel-wise classification is done by comparing the ground truth image and using results of HMRF- and GMM-based HMRF. The accuracy of a HMRF has the slight difference when compared to GMM-based HMRF. In the case of hard exudates, the HMRF achieves sensitivity of 66% and GMM-based HMRF achieves sensitivity of 56%. In case of soft exudates, the HMRF achieves sensitivity of 71% but the GMM-based HMRF achieves sensitivity of 43%. The HMRF achieves recall of 66% and GMM-based HMRF achieves recall of 55%. According to pixel-wise classification, the two methods are used to detect exudates but HMRF is more efficient than GMM-based HMRF from the observed results. In this paper the performance analysis of HMRF and GMM based HMRF are shown below:

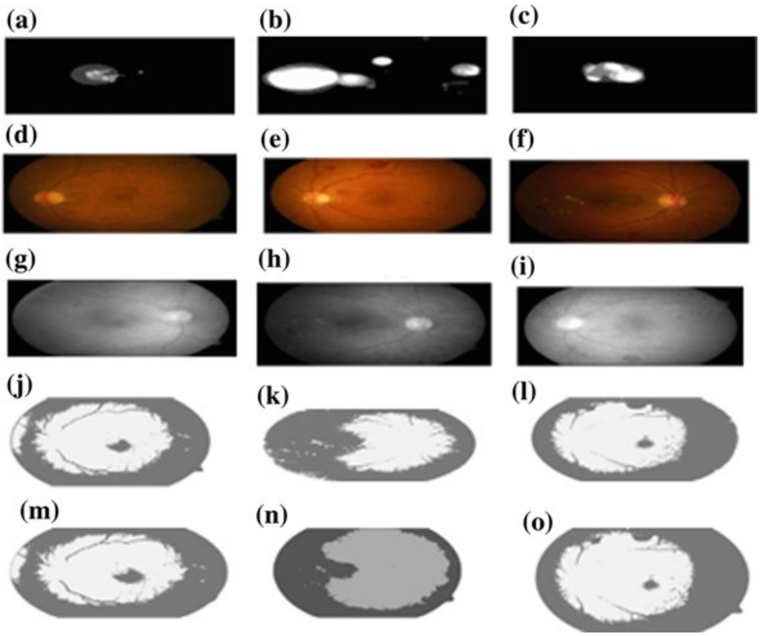


Fig. 3 **a**–**c** Ground truth images, **d**–**f** DIARETDB1, **g–i** closing operations, **j**–**l** initial labelling using k-means, **m**–**o** GMM-based HMRF

Table 1 Performance analysis of HMRF- and GMM-based HMRF

	Hidden Markov model		Gaussian mixture model based Hidden Markov model	
Pixel-wise classification	Average count of hard exudates	Average count of soft exudates	Average count of hard exudates	Average count of soft exudates
Accuracy	0.378026	0.376202	0.375152	0.373237
Sensitivity	0.664328	0.712173	0.56065	0.435036
Specificity	0.37393	0.375626	0.371404	0.393031
Precision	0.0176	0.001625	0.017016	0.001188
Recall	0.664328	0.712173	0.556559	0.435036

5 Conclusion

In this paper, we have done a comparative study of probabilistic models in colour fundus images to detect exudates. Automatic segmentation of exudates will help the clinicians to determine the disease level of diabetic retinopathy. The preprocessing technique is done using morphological operation with the standard dataset DIARETDB1. All the two methods are used to detect exudates, but HMRF is more efficient than GMM-based HMRF from the observed results.

References

1. Ram P, Swapna TR (2018) Profiling of hard exudates and microaneurysms to evaluate the progress of diabetic retinopathy. Int J Pure Appl Math 118(Special Issue 8):127–131
2. Ahmad A et al (2014) Image processing and classification in diabetic retinopathy: a review. In: 2014 5th European workshop on visual information processing (EUVIP). IEEE
3. Hussain F, Nandi Jaafar AK, Al-Nuaimy W (2010) Automated detection of exudates in retinal images using a splitand-merge algorithm. In: European signal processing conference, EUSIPCO
4. Choudhury S et al (2016) Fuzzy C means based feature extraction and classification of diabetic retinopathy using support vector machines. In: 2016 International conference on communication and signal processing (ICCSP). IEEE
5. Harini R, Sheela N (2016) Feature extraction and classification of retinal images for automated detection of diabetic retinopathy. In: 2016 Second international conference on cognitive computing and information processing (CCIP). IEEE
6. Priya R, Aruna P (2012) SVM and neural network based diagnosis of diabetic retinopathy. Int J Comput Appl 41(1)
7. Pires R et al (2014) Automatic diabetic retinopathy detection using BossaNova representation. In: 2014 36th Annual International Conference of the IEEE Engineering in medicine and biology society (EMBC). IEEE
8. Maher R, Kayte S, Bhable S, Kayte J (2015) Automated detection of diabetic retinopathy in fundus images. Inter J Emer Res Manag Tech 4(11):137–145
9. Kusakunniran W, Rattanachoosin J, Sutassananon K, Anekkitphanich P (2016) Automatic quality assessment and segmentation of diabetic retinopathy images. In: Region 10 conference (TENCON), 2016 IEEE, pp 997–1000
10. Shilpa BV, Nagabhushan TN (2015) An ensemble approach to detect exudates in digital fundus images. In: 2016 Second international conference on cognitive computing and information processing (CCIP). IEEE. Sing fuzzy C means. Int J Comput Appl 113(11)
11. Gandhi M, Dhanasekaran R (2013) Diagnosis of diabetic retinopathy using morphological process and SVM classifier. In: 2013 International conference on communications and signal processing (ICCSP). IEEE
12. Ganesh S, Basha AM (2015) Automated detection of diabetic retinopathy using retinal optical images. Int J Sci Technol Manage 4(2)
13. Sindhura A, Deva Kumar S, Ramakrishna Sajja V, Gnaneswara Rao N (2016) Identifying exudates from diabetic retinopathy images. In: 2016 International conference on advanced communication control and computing technologies (ICACCCT). IEEE, pp 132–136
14. Kaur G (2015) Review on: detection of diabetic retinopathy using SVM and MDA. Int J Comput Appl 117(20)
15. Deshmukh AV et al (2015) Features based classification of hard exudates in retinal images. In: 2015 International conference on advances in computing, communications and informatics (ICACCI). IEEE
16. Chowdhury AR, Banerjee S (2015) Detection of cotton wool spots from retinal images using fuzzy C means. Int J Comput Appl 113(11)
17. Rakshitha TR, Devaraj D, Prasanna Kumar SC (2016) Comparative study of imaging transforms on diabetic retinopathy images. In: IEEE international conference on recent trends in electronics, information and communication technology (RTEICT). IEEE, pp 118–122
18. Zhao Y et al (2017) Intensity and compactness enabled saliency estimation for leakage detection in diabetic and malarial retinopathy. IEEE Trans Med Imaging 36(1):51–63

19. Vijayan M, Jose R, Jose S, Abraham S, Joy J (2017) Study on quality of life assessment in diabetic retinopathy among patients with type 2 diabetes. Asian J Pharm Clin Res 10:116. https://doi.org/10.22159/ajpcr.2017.v10i7.18095
20. Roychowdhury, S, Koozekanani DD, Parhi KK (2014) Dream: diabetic retinopathy analysis using machine learning. IEEE J Biomed Health Inform 18(5):1717–1728
21. Kayte SN, Maher RS, Kayte CN (2015) Automated identification of hard exudates and cotton wool spots using biomedical image processing. Int J Comput Appl 131(5)

AUGEN: An Ocular Support for Visually Impaired Using Deep Learning

Reema K. Sans, Reenu Sara Joseph, Rekha Narayanan, Vandhana M. Prasad and Jisha James

Abstract Among the wide varieties of technologies, mobile phone technology has become popular and the usage of mobile phone applications is increasing day by day. Most of the modern mobiles are able to capture photographs. This can be used by the visually impaired to capture images of their surroundings which is then used to generate sentences that can be read out to the give visually impaired people a better knowledge of their surroundings. The content of an image is described automatically to them by which they can avoid seeking help from people around them. Computer vision is a field which can be used for gaining information from images or videos. The tasks which the human visual system can do can be done using computer vision. Visually impaired people can use these technologies in order to get better understanding of their surroundings.

1 Introduction

The problem in artificial intelligence which connects computer vision and natural language processing is describing automatically the contents of an image. To describe the contents of an image in proper English sentences is a challenging task. In this project, the system will be trained with different images. During training phase, the

R. K. Sans · R. S. Joseph · R. Narayanan · V. M. Prasad (✉) · J. James
Department of Computer Science & Engineering, Muthoot Institute of Technology & Science, Ernakulam, India
e-mail: vandhanasprasad1996@gmail.com

R. K. Sans
e-mail: reemaksans1996@gmail.com

R. S. Joseph
e-mail: reenusarajoseph@gmail.com

R. Narayanan
e-mail: rekhanarayanan96@gmail.com

J. James
e-mail: jishajames@mgits.ac.in

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_124

probability of producing the correct caption for the contents of the image should be maximized. During this training period, the system will be able to identify different objects contained in the image. How these objects relate to each other as well as their attributes and the activities they are involved in must be also known. This knowledge is then expressed in a meaningful sentence. Natural language processing is an application used for analysis and synthesis of natural language. The tasks such as automatic summarization, translation, speech recognition, etc. can be organized and structured by developers using NLP. NLP is characterized as a hard problem in computer science. The most difficult problems for artificial intelligence are processing of natural language. Many of the difficulties revolve around the issues of contexts. Neural networks are used in large areas, mostly to find the relationships between the data. They are used in traditional computer architecture to problems that the computer system cannot perform as good as the human visual system in areas like image recognition, making generalizations, that sort of thing. Researchers these days are constructing networks that are better at these problems. There are different neural networks used for training so as to perform natural language processing. It is a complex task to create a network manually, that will solve the difficult problems.

2 Literature Survey

Literature survey was done by referring papers related to the topic.

Paper [1] implemented the concept of Neural Image Caption (NIC) generator which can automatically view an image and generate a description in plain English. The key concept of NIC is Convolutional Neural Network (CNN) that encodes the image into a representation which is then followed by Recurrent Neural Network (RNN) which generates a correct caption for the image used as the input.

Paper [2] describes a system for caption generation which attempts to incorporate a form of attention with two variants, i.e., a hard attention mechanism and a soft attention mechanism. Since this system includes attention, it has the ability to visualize what the model sees. Standard backpropagation methods are used for training soft deterministic attention mechanism and hard stochastic attention mechanism trained by maximizing an approximate variation lower bound.

In paper [3], Recurrent Neural Network (RNN) is used to overcome the problem faced in spoken language understanding, which is semantic slot filling. A major task in spoken language understanding is to automatically extract semantic concepts.

In image sequence recognition, text recognition is one of the important tasks. Paper [4] describes a neural network architecture, in which feature extraction, sequence modeling, and transcription are integrated into a united framework. The proposed neural network model is named as Convolutional Recurrent Neural Network (CRNN), as it combines DCNN and RNN, and constructs an end-to-end system for sequence recognition.

Paper [5] deals with the various applications of Deep Belief Nets (DBN). In this study, DBNs to a natural language understanding problem is recognized. Natural

Language Processing (NLP) is a method by which the computers can analyze, derive, and understand meanings from the human language in a better and useful way.

Paper [6] talks about an extension of Long Short-Term Memory (LSTM) model. In particular, semantic information extracted from the image is added as extra input to each unit of the LSTM block. A Recurrent Neural Network (RNN) is a good choice to model temporal dynamics in sequences. Due to the issue of vanishing and exploding gradient, it is difficult for RNN to learn long-term dynamics.

Paper [7] depicts the mapping between the images and their corresponding description which is bidirectional. Recurrent neural network is then used to produce a visual representation of those images. The proposed system is capable of generating captions when an image is given as input as well as reconstructing visual features given in an image description.

Paper [8] uses the idea by which the humans perform task of recognizing visual sequences such as recognizing individual object, and then adding that recognized object to our internal representation of the sequence.

3 Motivation

In this era, where majority of the applications benefit the world of physically fit, there is a need to develop an application that would be a guide for the blind. Usually, they require the assistance of another individual to guide them which may create problems if the person is not trustworthy. There can be scenarios where the blind may be easily fooled. Considering all these issues into account, we decided to develop an application which favors the visually impaired. In this modern world, where technology is on its full swing in every field, there is a need for the blind to benefit from them as well. This application aims at providing the blind with a better idea of the things around them. Some of the few things that are been currently used to help visually impaired people to bear with their impairment and move on with their lives are braille, reading glasses, or a walking stick are just.

4 Proposed Method

Describing an image using properly formed sentences is a challenging task, but if successful will have a great impact. And, in this eon of technology demands a solution for this. Our project comes up with an android application that will lend a helping hand to the visually impaired people to understand their surroundings better. This application generates sentences for the images that have been captured by smartphone of the person in need, and speak out the caption being generated. Initially, the user captures the image using this application. After the image has been saved, it will be encoded into its feature set. This encoded image is then processed to generate

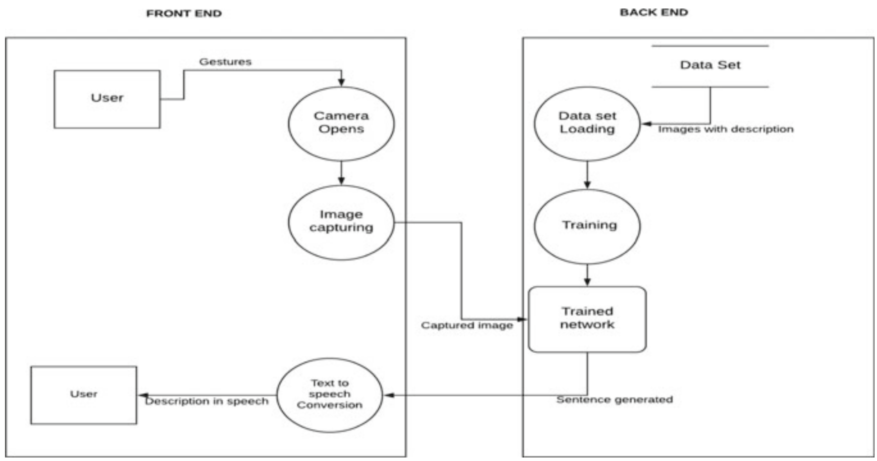


Fig. 1 System architecture

captions. The sentence so obtained is then converted to speech in order to favor the visually impaired. Therefore, the proposed system is as follows:

1. Network modeling:
   a. CNN encoding.
   b. RNN decoding.
2. Network Training:
   a. Dataset collection.
   b. Training the network.
3. Image capturing.
4. Generating captions using trained network.

Overall architecture of the proposed system (Fig. 1).

1. *Network modeling*

Network used is a pretrained Convolutional Neural Network (CNN)for image preprocessing. The image will be converted to a fixed vector representation. Then, Recurrent Neural Network (RNN) is used for generating image description. Specifically, Long Short-Term Memory (LSTM) network is used for framing description of the image. The core of the model includes an LSTM cell that processes one word at a time and which computes probabilities of the possible values for the next word in the sentence.

a. CNN encoding

In this CNN is used as an image encoder which will encode an image to its corresponding feature set for which it uses a pretrained. VGG16 model network with its pretrained weight is used as image encoder. Figure 2 shows architecture of VGG16 model. VGG neural network is a convolutional neural network for image classification. If an image is given as input to VGG16, it will find object name contained. It has 16 layers including convolutional layers, max-pooling layers, fully connected layers with 4096 nodes, and output layer with softmax as activation function with 1000 nodes. In this implementation, last softmax layer of VGG16 is removed and the vector of dimension (4096) is obtained from the second last layer. Figure 2 shows layers of VGG16.

b. RNN decoding

Generation of caption is achieved using an LSTM network. At the current time step, the network takes the image vector along with the partial captions and generates the word that is having the highest probability as output.

$$\theta^* = \text{argmax}_\theta \sum_{(I,S)} \log\ p(S \vee I; \theta) \tag{1}$$

Equation (1) is used to maximize the probability of generating correct description, where θ represents the parameters of the model, I represent image, and S represent sentence.

$$p(S|I) = \sum_{i=0}^{N} \log p(S_t|I, S_0, \ldots, S_{t-1}) \tag{2}$$

To join the probability of N sentences $S_0, S_1, \ldots, S_N$ chain rule is applied to the model as per Eq. (2).

2. *Network Training*

a. Dataset Collection

For the training and testing of network, Flickr8k dataset has been used. Flickr8k consists of 8000 images out of which 6000 are used for training, 1000 are used for testing, and remaining 1000 are used for validation.

b. Training the network

CNN and RNN are used for generating caption. For this, the network is trained using the downloaded dataset. To speed up training, each image was pre-encoded to its corresponding feature set. One hot encoding of the words is not preferred as each caption may consist of large number of unique words. Instead, an embedding model is trained that takes a word and outputs an embedding vector of dimension

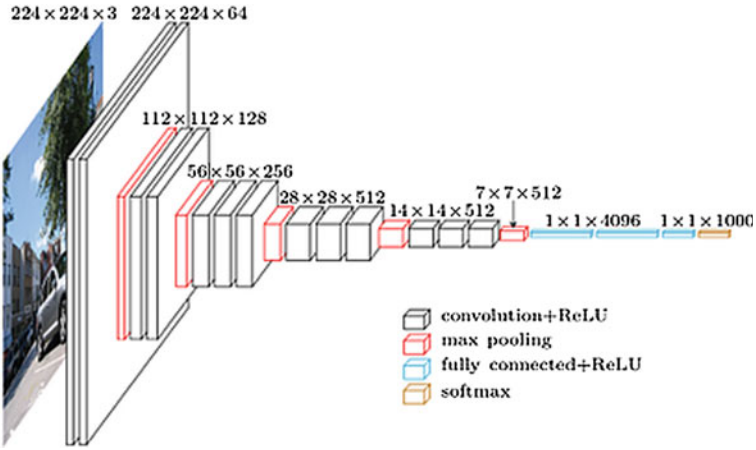


Fig. 2 VGG16 layers

(1, 128). LSTM-based network is used for generating caption. During this training phase, network is being taught how to generate caption for the image by analyzing the dataset provided. These networks were trained with 50 epochs. Once the training of the network is completed, a weight model file is created. This file will contain the learned weights of network during training. Test image which is in vector format is then given as input to the weight model for formation of caption.

3. *Image capturing*

This module is about activating the application using a gesture. Application is activated by shaking the mobile phone, as a result of which camera activity will be invoked and the user can capture the image. The images captured by the user will be sent to the server which will be stored in a file for processing. Then, captions will be generated for the same.

4. *Generating captions using trained network and converting to speech*

Caption is generated for the saved image using trained network. The image from mobile is given as input to trained network which will generate caption for the same which is saved to a file. Then, this caption is sent back to mobile phone. The caption so obtained is then converted to speech in order to favor the visually impaired.

5 Result

Our project aims to generate captions for the image given as input. For performing this, plenty of image data is required. We have access to various image datasets such as Flickr30k, MS COCO, SBU, Pascal, Flickr8k, etc. This model is trained with

Fig. 3 Input 1

Fig. 4 Input 2

Flickr8K dataset. We obtained a training accuracy of 78%. The model generates three captions for each image from which the caption with highest probability is considered.

Following figures were given as input:

The captions generated for Fig. 3 are as follows:

1. A person rows a boat over a large body of water.
2. A person riding a surfboard on a body of water.
3. A person riding on a surfboard in the water.

The captions generated for Fig. 4 are as follows:

1. A cat sits alone in dry grass.
2. Couple of kittens is standing in a field.
3. A group of chicken is standing in a field.

6 Future Work

The proposed system focuses on producing caption for images. This would work more efficiently if extended to generate the caption from a video. Some of the issues like blurred images, out of focus, etc. associated with the captured images can be resolved on using video as the input. A faster and an efficient network will help to overcome the delay required in generating captions for the image being given as input to the server.

7 Conclusion

This paper represents a developed user-friendly android application that will certainly lend a helping hand for the visionless. Most of the applications are developed to benefit the majority, i.e., the physically fit. But, there is also a need to develop an application that will make the life of the minority much easier. Usage of mobile phone applications is increasing day by day. And, nowadays there are technologies which have the potential to significantly emend the lives of visually impaired ones, which will provide them a better knowledge of the surroundings for the blind. The main inspiration for our work is the recent advancement in machine translation. AUGEN, an android application which generates meaningful sentence for the image that has been captured by the smartphone of the person in need and speak out the sentence formed to favor the visually impaired. The possibilities of using technology for making the life better each day are never ending and we are just beginning to scratch the surface.

References

1. Vinyals O, Toshev A, Bengio S, Erhan D (2015) Show and tell: a neural image caption generation. In: Computer vision and pattern recognition (CVPR), IEEE conference
2. Xu K, Ba JL, Kiros R, Cho K, Courville A, Salakhutdinov R, Zemel RS, Bengio Y (2015) Show, attend and tell: neural image caption generation with visual attention. In: International conference on machine learning, pp 2048–2057
3. Mesnil G, Dauphin Y, Yao K, Bengio Y, Deng L, Hakkani-Tur D, He X, Heck L, Tur G, Yu D, Zweig G (2015) Using recurrent neural networks for slot filling in spoken language understanding. IEEE/ACM Trans Audio Speech Lang Process 23(3):530–539
4. Shi B, Bai X, Yao C (2016) An end-to-end trainable neural network for image-based sequence recognition and its application to scene text recognition. IEEE Trans Pattern Anal Mach Intell 39:2298–2304
5. Sarikaya R, Hinton GE, Deoras A (2014) Application of deep belief networks for natural language understanding. IEEE/ACM Trans Audio Speech Lang Process 22(4):778–784
6. Leuven XJKU, Gavves E, Fernando B, Tuytelaars T (2015) Guiding the long-short term memory model for image caption generation. In: The IEEE international conference on computer vision (ICCV), pp 2407–2415

7. Chen X, Lawrence Zitnick C (2015) MindsEye: a recurrent visual representation for image caption generation. In: The IEEE conference on computer vision and pattern recognition (CVPR), pp 2422–2431
8. Ba JL, Mnih V, Kavukcuoglu K-R (2014) Multiple object recognition with visual attention. arXiv: 1412.7755 [cs.LG]

A Novel Flight Controller Interface for Vision-Guided Autonomous Drone

R. Senthilnathan, Niket Ahuja, G. Vyomkesh Bhargav, Devansh Ahuja and Adish Bagi

Abstract Aerial vehicles are rapidly exploring places in a variety of applications in the service sector such as transport assistance and other logistics involved sectors. Most applications require local perception for the aerial vehicles wherein vision system is the most powerful information. Such local perception may be designed based on landmarks on the ground below. This paper details the work where a computer vision system is aiding in landmark identification, namely, a line strip on the ground below. The vision system is implemented in a real-time controller, and a novel and simple interface solution is presented to interface the vision controller to the flight controller. The aerial vehicle under consideration is quadrotor type. The paper presents the details of the image processing algorithm along with the hardware details of the flight control interface. The vision system is developed to identify line strips which share sufficient contrast with the background under daylight condition. The image analysis is performed to continuously extract the lateral position and yaw error. Such an interface would serve as a computationally cheap and easy solution over traditional programmable flight control interface solutions such as MAVLINK.

1 Introduction

Drones are aerial vehicles which are capable of performing complex and hazardous operations, acquire data from the ground, process the data and control the flight of the drone over the desired course [1, 2]. The conventional drones are manually controlled via a remote control. An autonomous drone is the one which is controlled by a controller on-board. The flight of an autonomous drone is secured without any human intervention. The autonomy of the drone is supported by real-time image or video processing, obstacle avoidance, line following, etc., applications. The raw data is then processed and fed to the on-board flight controller. The stability of the drone in motion is generally established through the calibrated flight controller.

R. Senthilnathan (✉) · N. Ahuja · G. Vyomkesh Bhargav · D. Ahuja · A. Bagi
SRM Institute of Science and Technology, Kattankulathur, India
e-mail: senthilnathan.r@ktr.srmuniv.ac.in

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_125

Generally, accurate calibrations such as accelerometer, speed controllers and motor testing need to be performed. Manoeuvrability in confined spaces is better achieved with the application of autonomous drones over the conventional drones which are remotely controlled. In the work presented in this paper, a quadrotor drone is utilized. Quadrotor aerial vehicle is a type of drone that consists of four rotors and two pairs of counter-rotating, fixed-pitch blades located at four corners of the body [3]. The design is simple in its mechanical structure and hence the dynamics. Furthermore, the use of four rotors allows each individual rotor to have a smaller diameter. In this way, the damage that may be caused by the rotors in case of undesirable situations is reduced [3].

Recent work reported in the literature have reported many ways of enhancement of visual perception, visual processing, redundant systems for landing and manoeuvring in structured and unstructured environments. Different algorithms have been proposed and developed to address current challenges [3–7]. A quadrotor which uses on-board processing for localization and mapping with stereo vision to manoeuvre it in indoor and outdoor GPS denied environment is presented in [5]. Although interstitial/GPS approach was used for the quadcopter to perceive the environment, it suffers from two drawbacks: necessity to receive GPS signals and lack of precision in position [7].

Vision-based aerial vehicles provide complete information of the environment in which they are located with the ability to live stream the information of their surroundings, and in the recent past, this has helped the user to manoeuvre the aerial vehicle as per their need, operated from a remote location. Vision sensor could also be used to self-estimate the attitude and altitude in correspondence with other inertial measurements [8]. The pseudo roll and pitch can be decided from the on-board video or image streams [9]. Experiments on only vision-based navigation and obstacle avoidance have been achieved on small rotary wing aerial vehicles [10]. The vision system itself faces a lot of challenges which can be due to the camera, surrounding environment or to a certain extent the aerial vehicle itself due to its body dynamics. Previous work [11] by the author involving vision guidance for drone docking involves a vision in the loop feedback apart from the motion control system realized through the flight controller. A similar approach is adopted for the current work where a cascaded control approach between the vision controller and the flight controller. The major savings is the computational cost involved in the software generation of the pulse position modulation (PPM) signals which are required for the flight controller. The current work is based on a state-of-the-art hardware PPM converter and high-speed update of PWM generation in a dedicated controller.

2 Drone Specifications

The quadrotor used for the experiment is a controlled in plus-configuration. It is controlled through a primary flight controller, namely, Pixhawk PX4. The various specifications of the components of the flight controller are listed in Table 1. The

Table 1 Specifications of the quadcopter

Specifications	Parameter
Frame	F450 Nylon fibre frame
Flight controller	Pixhawk PX4
Motors	2 × CW motors, 2 × CCW motors
ESC	30 A
Vision controller	Nvidia Jetson TK1
Camera	3 mm focal length lens
Battery	3000 mAh Li–Po
Propellers	2 × CW, 2 × CCW
Payload	1.2 kg
KV rating of BLDC motor	1000 kV
No load current	10 V: 0.5 A
Current capacity of battery	12 A/60 s
No of cells	2–3 Li–Po
Motor size	27.5 mm × 30 mm
Shaft diameter	3.175 mm
Minimum ESC specification	18 A
Type of propeller	1045 propeller
Diameter	10 in.
Pitch	4.5 in.
Propeller diameter	254 mm
Centre bore diameter	6 mm front and 9 mm reverse side
Centre seat TH	6 mm
Voltage of the motor	11.1 V
Max continuous current	30 C(90.0 A)
Input voltage	DC 6–16.8 V
Running current	30 A
Size of the motor	26 mm × 23 mm × 11 mm (L × W × H)

photograph of the fabricated quadcopter on a spherical gimbal is shown in Fig. 1. The vision algorithm is executed in a state-of-the-art embedded controller, namely, Jetson TK1 from Nvidia Corporation. The controller has an integrated Tegra processor consisting of a quad-core ARM CPU and 192 core Kepler GPU. The controller runs Linux Ubuntu version called Linux for Tegra custom tailored to suit the demands of embedded applications. The controller has GPIO which operates with logical high at 1.8 V which in some sense is a drawback since most electronics used in motion control operates on 5 V TTL logic. This requires a level translation from 1.8 to 5 V and vice versa. The proposed drone control strategy involves a standard flight controller since the robustness of commercially available flight controllers is exceptionally good. The key specifications are listed in Table 1.

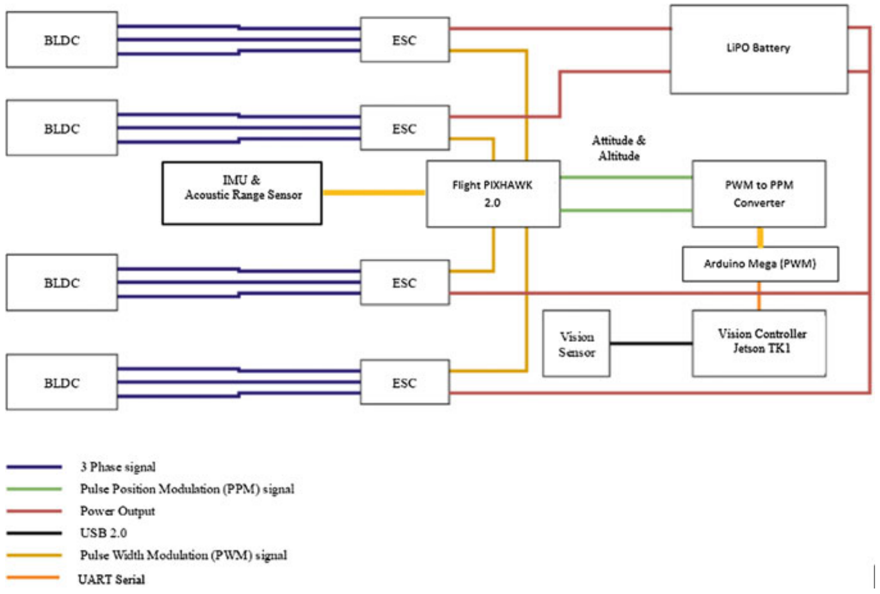


Fig. 1 Hardware Interface

In order to interface the flight controller to the vision controller, two main solutions are available. The first solution is through the MAVLINK protocol of serial commands to send and receive data from the flight controller. The second method which is adopted in the current work is to hack into the RC receiving channels of the flight controller and mimic the PPM signals for various attitude and altitude controls as it would be applied from the remote radio controller. In order to accomplish the latter method, PPM signals corresponding to pitch, roll, yaw and throttle must be generated by the vision controller which causes the flight controller to generate suitable PWM signals which get applied to the electronic speed controllers. Figure 1 illustrates the hardware interface strategy adopted in the current work.

3 Imaging Hardware

The current work uses equipment that is necessary for successful implementation of vision guidance system which is described as follows.

3.1 Scene Description

The current work is proposed for indoor environment with an area of approximately 4000 ft^2 and a height of approximately 50 ft from ground to roof. The line to be followed by the vehicle is made using a light coloured masking tape of sufficient width stuck on the ground of dark complexion. The closed enclosure ensures no wind turbulence being offered to the vehicle. It also ensures ambient daylight to enter the enclosure.

3.2 Camera

The camera helps to broadcast live image to the system. The work presented in the paper demonstrates pre-acquired images from the camera which acts as the input data to the simulated model of the quadrotor vehicle. The images are processed in order to obtain the information necessary to decide the movement of the vehicle. The camera is mounted underneath the quadcopter through a mounting, and hence, image acquisition mimics the similar acquisition pose for the camera. The camera selected is of 1.3 megapixel. The camera consists of an Aptina sensor with USB 2.0 interface. It also has a feature of day and night low light vision for environmental monitoring camera system. It has 2.1 mm wide angle lens for all kinds of camera control system. It weighs around 90 g.

3.3 Camera Modelling and Image Acquisition

The camera modelling process relates the metric real-world information to information in image space. For the current work, a scaled orthographic projection type of camera model would suit the purpose since the line strip is thoroughly planar. If a point $P(X, Y, Z)$ on the real world with reference to the camera is mapped to the point $I(x, y)$ on the image plane then the relation as per a scale orthographic projection may be defined as

$$x = s.X \; \& \; y = s.Y$$

where s is the scaling factor which is a function of the standoff distance, Z. The image acquisition adopted for acquiring the strip image sequence is asynchronous in nature. In this method, the image is first completely captured whenever a start command is read by the system. This acquisition of image takes place for a fixed time after which the capturing operation stops.

4 Vision Algorithm

The basic objective of the vision system is to obtain the error along the horizontal axis (controlled by roll motion) of the image measured from the image centre reference to the strip centre and ensure that the axis of the strip in the field of view of the camera is along the vertical axis of the image (controlled by pitch motion).

4.1 Image Processing

The current work uses on-board vision system to acquire the image of the scene which is processed by the on-board vision controller. The image captured by the camera is subjected to the vision algorithm which extracts line features from it. The first step is to convert the RGB colour image to grayscale image (Fig. 2).

The grayscale image is converted to binary image based on a predetermined threshold. The threshold intensity will make sure that the pixels having intensity less than threshold will be converted to black and the pixels having intensity greater than the threshold will be converted to white pixels. The threshold value depends upon the illumination of the scene in consideration. This makes the strip to be completely white on a black background which makes the extraction of line features conveniently. Random noise in the binarization is eliminated using area granulometry since the strip is assumed to be the largest region in the binary image. Inconsistencies in colour of the strip result in presence of holes which are suitably eliminated using morphological closing (Fig. 3).

After computing the aspect ratio and determining the orientation of the strip, its width in pixels is obtained and its mean is computed, and a skeletonized strip which is of one-pixel width is obtained.

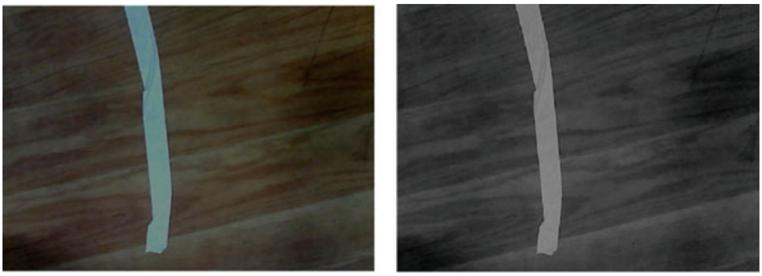


Fig. 2 RGB colour image and the corresponding monochrome version

Fig. 3 Binary image before and after noise removal and skeletonized version

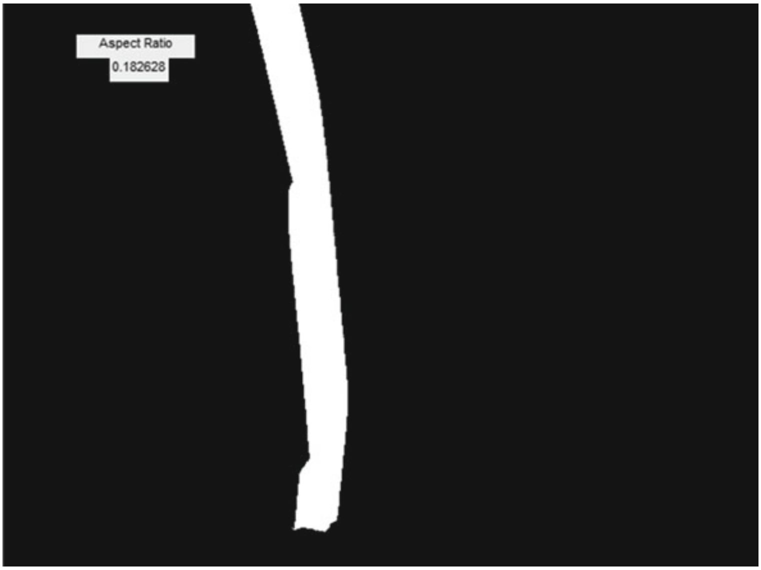


Fig. 4 Aspect ratio of the strip

4.2 Image Analysis

The vision algorithm takes input as image and carries out various operations on it to give numerical data as output at necessary stages of operation. The aspect ratio of the strip is obtained to determine the orientation of the strip in the image frame. Figure 4 illustrates the computation of the aspect ratio.

A skeleton line is obtained by processing the image. The length of this line is calculated via image analysis. The measurement of the skeletal line depends on any arbitrary orientation of the line, in the image frame captured by the camera. Either, a combination of row and column coordinates or just the row and column coordinates are taken into consideration to calculate the length of the line. After getting the total length, the centre of this line is computed by taking the mean of the corresponding coordinates. The centre of the image is obtained by taking the mean of the row and column coordinates of the complete image. A window is generated enclosing the processed image. A quarter length of the line is taken on either side from its

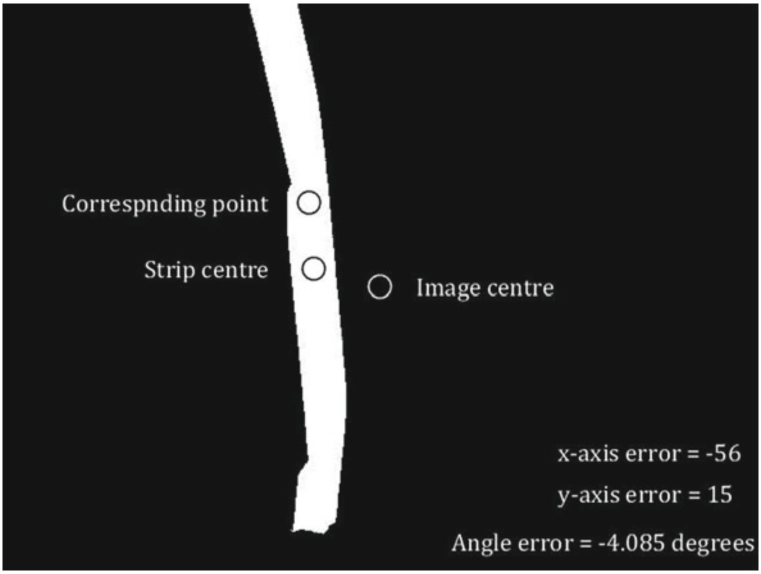


Fig. 5 Corresponding point and other results

centre, to create the window. The window intersects the line at two points. The point lying in-line with the heading direction of the quadrotor aerial vehicle is taken into consideration. The error angle between the line joining the centre point of the line and the corresponding point in the present frame and the previous frame is computed. The error between the line centre and the image centre is computed. These two errors are corrected by the flight controller to get the desired motion of the line following vehicle (Fig. 5).

5 Vision Guidance

Basically, three errors are desired to be corrected based on the information obtained from the vision system, namely, x, y, heading direction. Let dx and dy give the difference in the centroid positions which is calculated between the image centre and the strip centroid. Line following of quadrotor refers to controlling of position $P_{\mathrm{E}}(x, y)$ and angular position, i.e. pitch and roll. The initial condition of the quadrotor is such that the strip is in the field of view of vision sensor. The angular positions, pitch(ϕ) and roll(θ) are controlled by the outer loop controlling positional error through vision. The altitude of the quadrotor is maintained through a separate control loop for maintaining the range from the ground below is considered with a range sensor in the loop. The quadrotor starts after ensuring the initial marker condition which in the current work is

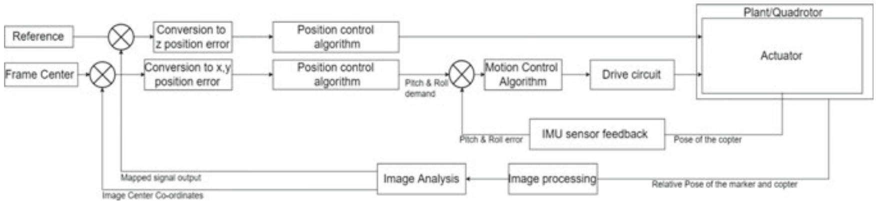


Fig. 6 Vision in the loop control block diagram

a discontinuous line strip. The same fiducial is also applicable to the ending condition. The block diagram in Fig. 6 illustrates the quadrotor control strategy adopted.

5.1 X-Y Position Control

The position control loop depends on the instantaneous position error $P_{\mathrm{E}}(x, y)$ which is calculated by keeping the vision in loop so that the error signal is generated in the form of pixels. The control law is completely planned in image space. The roll and pitch angles are generated for any given error in x and y axes of the image. The position controller generates output that is further fed to control the pitch and roll of vehicle.

5.2 Attitude Control

Motion control loop is the inner loop of the system which is responsible for controlling the pitch and roll of quadrotor. As per Fig. 6, the pose-related information is given as the feed to the loop through inertial motion unit (IMU). The pitch and roll signal demand signals generated through the outer X, Y position control loop is compared to the feed signals from the IMU. The generated error signal is fed to the motion controller. The output of motion controller is then converted to motor signals through drive circuit and is used for actuation of rotors in the system.

5.3 Altitude Control

Z position control loop is required to maintain the altitude of the quadcopter. The stable altitude would ensure no magnification changes during the manoeuvring of the quadcopter. The flight controller has a default altitude measurement device which is based on atmospheric pressure. This sensor performs poorly in terms of the resolution indoor conditions; a redundant acoustic ranging sensor compatible with flight controller is used. The sensor is suitably configured in the flight controller planning software to act as replacement to the barometric sensor.

Thus, the paper presents a novel interface solution for vision-guided aerial vehicles which can be very simple in terms of the construction and principle and yet can be an effective alternative to complex but the current benchmark for flight controller communication, namely, the MAVLINK protocol.

References

1. Puri A (2005) A survey of unmanned aerial vehicles (UAV) for traffic surveillance. Department of Computer Science and Engineering, University of South Florida
2. How Jonathan P (2008) Real-time indoor autonomous vehicle test environment. Control Syst IEEE 28(2):51–64
3. Bi Y, Duan H (2013) Implementation of autonomous visual tracking and landing for a low-cost quadrotor. Optik-Int J Light Electron Opt 124(18):3296–3300
4. Meier L, Tanskanen P, Heng L, Lee GH, Fraundorfer F, Pollefeys M (2011) PIXHAWK: a system for autonomous flight using onboard computer vision. In: 2011 IEEE international conference on Robotics and automation (ICRA). IEEE
5. Chae H, Park J, Song H, Kim Y, Jeong H (2015) The IoT based automate landing system of a drone for the round-the-clock surveillance solution. In: IEEE international conference on advanced intelligent mechatronics (AIM). IEEE
6. Pattersona T, McCleana S, Morrowa P, Parra G, Luob C (2014) Timely autonomous identification of UAV safe landing zones. Image Vis Comput 32(9):568–578
7. Blösch M (2010) Vision based MAV navigation in unknown and unstructured environments. In: 2010 IEEE international conference on robotics and automation (ICRA). IEEE
8. Roberts PJ, Walker RA, O'Shea P (2005) Fixed wing UAV navigation and control through integrated GNSS and vision. In: AIAA guidance, navigation, and control conference and exhibit
9. Dusha D, Boles WW, Walker R (2007) Fixed-wing attitude estimation using computer vision based horizon detection, pp 1–19
10. Calise AJ (2003) Applications of adaptive neural-network control to unmanned aerial vehicles. In: AIAA/ICAS international air and space symposium and exposition: the next, vol 100
11. Senthilnathan R, Nikhil S, Sasikala R, Maruvada A (2017) Model based design of vision guided vertical landing system for quadrotor. Int J Control Theor Appl 10(9):803–819

Sliding Discrete Fourier Transform for 2D Signal Processing

Anita Kuchan, D. J. Tuptewar, Sayed Shoaib Anwar and Sachin P. Bandewar

Abstract Discrete Fourier Transform (DFT) is the most frequently used method to determine the frequency contents of the digital signals. As DFT will take more time to implement, this paper gives the algorithm for the fast implementation of the DFT on the Two-Dimensional (2D) sliding windows. To fast implement DFT on the 2D sliding window, a 2D DFT (here 2D SDFT) algorithm is stated. The algorithm of the proposed 2D SDFT tries to compute current window's DFT bins directly. It makes use of precalculated bins of earlier window. For a 2D input signal, sliding transform is being accelerated with the help of the proposed algorithm. The computational requirement of the said algorithm is found to be lowest among the existing ones. The output of discrete Fourier transform and sliding discrete Fourier transform algorithm at all pixel positions is observed to be mathematically equivalent

1 Introduction

In several applications of image processing, frequency domain offers an improvement over performing a similar job in the time domain. At times the improvement is just simpler or more hypothetical algorithm. Often the largest obscurity in working in frequency domain is concern with calculation of Fast Fourier Transform [1]. If the frequency-domain data must be reorganized constantly in a real-time application, the difficulty and latency of the FFT can become a significant obstruction to achieve system goals and keeping cost and power consumption low.

A. Kuchan (✉) · D. J. Tuptewar · S. S. Anwar · S. P. Bandewar
Mahatma Gandhi Missions College of Engineering, Nanded, India
e-mail: anitakuchan@gmail.com

D. J. Tuptewar
e-mail: tuptewar_dj@mgmcen.ac.in

S. S. Anwar
e-mail: sayed_shoaib@mgmcen.ac.in

S. P. Bandewar
e-mail: bandewar_sp@mgmcen.ac.in

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_126

For detection and processing signals, there are a number of real-time applications, such as touchscreen sensing, medical imaging and communication systems, and radar uses frequency-domain algorithms. There are many such applications where complexity should be low while reducing the latency. Sliding DFT method [2–4] gives frequency-domain information on the sample by sample basis with considerably less computations than the FFT for each update [5]. Various transform codings constitute an important module of existing image processing applications. In the sliding transform, the window is shifted by one sample ahead at a time; this procedure is repeated throughout the image. There are various algorithms like DFT for moving window [6], Walsh Hadamard Transform (WHT) [7], Running Walsh Hadamard Transform (RWHT) [8], Gray-Code Filter Kernels (GCK) [9], WHT on sliding windows [10, 11], Hopping DFT (HDFT) [12] and Sliding WHT (SWHT) [13] which works on sliding windows. For the implementation of sliding DFT, the discrete orthogonal transform becomes more complex. Various algorithms are developed in the last couple of years for the rapid calculations of the sliding orthogonal transforms. Here, sliding DFT is introduced.

The rest of the paper includes the necessary background and the key features of the proposed scheme. Also, the performance evaluation experiments are presented in the subsequent phases of the script. Remaining part presents the results and finally the conclusion.

2 Proposed Method

Considering a signal $y(n)$, which is represented as below:

$$\ldots, y(-2), y(-1), y(0), y(1), y(2), \ldots y(n), y(n+1), y(n+2) \ldots.$$

In various applications of signal processing, one might desire to study the signal with the help of sliding window of fixed size. Assuming at the time instant n, a windowed series $y(n)$ containing N number of samples is given by Eq. (1)

$$y(n) = \{y(n-N+1), y(n-N+2), \ldots, y(n-2), y(n-1), y(n)\} \quad (1)$$

after taking DFT of given sequence y(n), we get $Y(n)$ which is represented as

$$Y(n) = \{Y_0, Y_1, Y_2, Y_3 \ldots Y_{N-3}, Y_{N-2}, Y_{N-1}\} \quad (2)$$

That means, to compute DFT sequence $Y(n)$, we have to perform DFT at each and every time instant n. If it is needed, then one may use FFT which is fast and computationally efficient. But this is not the scenario always, in some applications, the requirement is of only kth value of the DFT, consider Y_{k-1}, then Fast Fourier Transform (FFT) is not computationally efficient. Here with the help of FFT, we are considering complete DFT sequence and to ignore redundant values, is there any solution? The answer for this is, there we have a computationally efficient algorithm, which is known as Goertzel algorithm. This Goertzel algorithm [14] helps to get one-off Y. Similarly, when we are taking into consideration a sliding window, Sliding DFT (SDFT) is most efficient. The concept of SDFT introduced from the two succeeding time instants [11]. Consider two succeeding time instants as $n-1$, n and $y(n-1)$, $y(n)$ as windowed sequences which contains basically equal elements.

For example, let the window length of $N = 8$, and then at $n = 7$, the sequence is given as

$$y(7) = \{y(0), y(1), \ldots, y(6), y(7)\} \tag{3}$$

$Y(7)$ is calculated with the help of 8-DFT. And at $n = 8$, we get

$$y(8) = \{y(1), y(2), \ldots., y(7), y(8)\} \tag{4}$$

here we observe remarkable similarities between $y(n-1)$ and $y(n)$ in the SDFT for efficient calculations.

Prior to introducing the Sliding DFT, we memorize one of the basic important properties of DFT which is circular shift property. If N-DFT of a sequence $y[n]$ is given by $Y[k]$, then we can write as

$$y[((n-m))N] \leftrightarrow W_N^{km} Y[k] \tag{5}$$

where $W_N = \mathrm{e}^{\frac{-j2\pi}{N}}$; thus, the above equation gives circularly shifted series of $y[n]$.

Specifically, if we use circular shift property, then the sequence is circularly moved by one time instant (in the left direction), and then, the DFT value Y_k will be

$$Y_k \rightarrow Y_k \mathrm{e}^{j2\pi k/N} \tag{6}$$

apart from importance of N, one can get $y(n)$ from $y(n-1)$ with the help of following steps:

1. Restore the foremost element of sequence $y(n-1)$ with the very last element of $y(n)$, e.g. initiate with (3) and exchange y_0 by y_8.
2. Apply circular shift property (which will shift sequence to the left).

These two steps show the way to easy things in the DFT domain, which is represented as

$$Y_k = \sum_{n=0}^{N-1} y(n) W_N^{nk} = y(0) + y(1)W_N^k + \cdots + y(N-1)W_N^{(N-1)k} \tag{7}$$

From Eq. (7), one can examine that the first component $y(0)$ is remained unchanged in the DFT formula, and now, apply step1 to have an in-between sequence

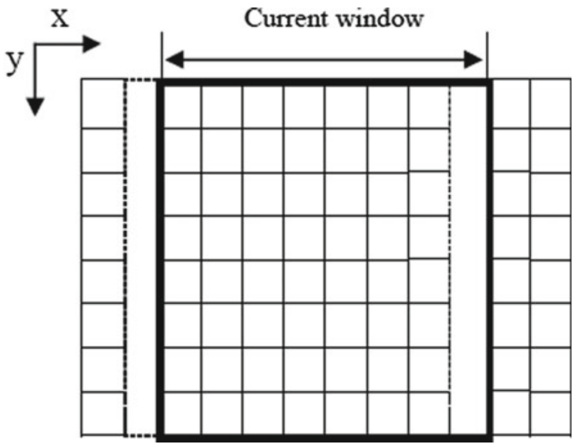


Fig. 1 Window of size 8 × 8

$$\widehat{y(7)} = \{y(8), y(1), \ldots, y(6), y(7)\} \tag{8}$$

we can write from (7)

$$\widehat{Y(7)} = Y_k(7) - y(0) + y(8) \tag{9}$$

now the DFT of $y(8)$ is denoted by $Y_k(8)$ then by applying property we get

$$Y_k(8) = \widehat{Y_k(7)}\, \mathrm{e}^{j2\pi k/N} \tag{10}$$

Hence, the relation between DFTs of two consecutive windowed sequences $y(n-1)$ and $y(n)$, having length N, is represented as

$$Y_k(n) = [Y_k(n-1) - y(n-N) + y(n)]\mathrm{e}^{j2\pi k/N} \tag{11}$$

Equation (11) represents the basic concept of Sliding Discrete Fourier Transform (SDFT). Only one complex multiplication and two real additions per output sample are required to implement SDFT. However, because of the recursive nature of SDFT, output must be calculated for each and every new input sample. We are applying this algorithm to 2D signals that is image. Figure 1 shows window of size 8 × 8 graphically.

Considered R_{M} signifies the numbers of real multiplications and R_{A} represents real additions. Then, the total analysis of the 2D SDFT method is summarized as

$$R_{\mathrm{M}} = 4\left(M^2 + M\right) \tag{12}$$

and

$$R_{\mathrm{A}} = 4M^2 + 3M + 2 \tag{13}$$

Table 1 Computational requirement of DFT, FFT and SDFT

Algorithm	Computational requirement	Window size $M \times M$
DFT	R_{M}	$8M^3$
	R_{A}	$8M^3 - 4M^2$
FFT	R_{M}	$4M^2 \log_2 M$
	R_{A}	$6M^2 \log_2 M$
2D-SDFT	R_{M}	$4(M^2 + M)$
	R_{A}	$4M^2 + 3M + 2$

Table 2 Computational requirement of DFT bins for different window sizes

Algorithm	Operation	Number of computed DFT bins			
		4 × 4	8 × 8	12 × 12	16 × 16
DFT	R_{M}	512	4096	13,824	32,768
	R_{A}	448	3840	13,248	31,744
FFT	R_{M}	128	768	2065	4096
	R_{A}	192	1152	3098	6144
2D-SDFT	R_{M}	80	288	624	1088
	R_{A}	78	282	614	1074

A computational evaluation is summarized in Table 1 where the computational analysis of 2D SDFT, FFT and DFT is shown. After comparing all these algorithms, it is found that 2D sliding DFT method is having less computational complexity in comparison with available methods. Consequently, for some applications, we suggest the use of the 2D sliding DFT method where the computational complexity is a significant concern.

This 2D SDFT algorithm calculates DFT points separately that is one sample after another. This indicates, whenever one wants to calculate DFT at few pixels, the algorithm has to implement the DFT at all time instants. In many image processing applications, all DFT bins need not to be calculated. We summarize the computational requirements of the algorithms when only some part of the DFT bins are calculated, and it is shown in Table 2.

Table 2 shows that as the number of estimated DFT bins reduces, the computational requirements of all algorithms decrease. From Table 2, we can say that the performance of the SDFT method gives better improvement when the number of DFT bins to be calculated is decreased.

The computational analysis is done for the window sizes 4 × 4, 8 × 8, 12 × 12 and 16 × 16. Graphically, it is represented in Fig. 2.

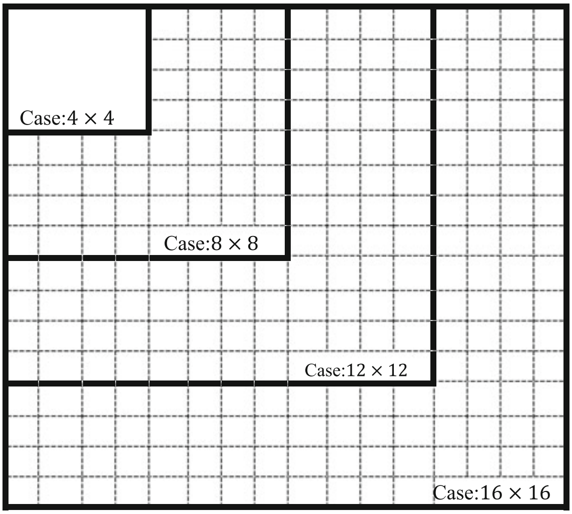


Fig. 2 DFT bins of window size 4 × 4, 8 × 8, 12 × 12 and 16 × 16

3 Experimental Results

We executed this algorithm using several CIF, 4CIF and QCIF videos with the rate of 30 frames per second. Sliding transform was applied considering the first 100 frames of each video sequence.

Mathematically if we compare algorithms like DFT, FFT and SDFT, all these algorithms give the same transform results. The difference is there in computational analysis. Initially, we compute the time required for every algorithm when DFT bins are to be calculated. The experimental results observed by giving various CIF and 4CIF sequences as input are presented in Tables 3 and 4, respectively.

Here, we denote $\Delta T\%$ the average processing time reduction of the existing 2D SDFT as compared to SDFT. This average processing time reduction is calculated by using the formula given below:

Table 3 Experimental results with comparison for the CIF videos

Sequences	DFT	FFT	Existing 2D SDFT	SDFT	$\Delta T(\%)$
	Time (ms)	Time (ms)	Time (ms)	Time (ms)	
City	534,940.06	69,685.00	19,875.65	14,962.73	24.71829
Soccer	534,823.75	69,685.29	19,973.48	19,401.124	2.86558
Harbour	534,716.63	69,689.14	19,836.59	18,944.464	4.497376
Average	534,826.8133	69,686.47,667	19,895.24	17,769.43,933	10.69375

Table 4 Experimental results with comparison for the 4CIF videos

Sequences	DFT	FFT	2-D SDFT	SDFT	$\Delta T(\%)$
	Time (ms)	Time (ms)	Time (ms)	Time (ms)	
City	2,256,673.25	353,330.72	83,643.72	51,640.475	38.26138
Soccer	2,256,662.25	354,086.63	83,701.13	52,137.789	37.70958
Harbour	2,258,794.18	353,240.59	83,578.79	53,547.582	35.93161
Average	2,257,376.56	353,552.6467	83,641.21333	52,441.94867	37.30086

Table 5 Experimental results for the QCIF videos

Sequences	SDFT
	Time (ms)
City	12,594.692
Soccer	11,767.736
Harbour	12,432.937
Average	12,265.121

$$\Delta T(\%) = \frac{T_{\text{algorithm}} - T_{\text{Proposed}}}{T_{\text{algorithm}}} \times 100$$

where T_{Proposed} indicates the calculated processing time of the SDFT and $T_{\text{algorithm}}$ indicates the calculated time of 2D SDFT.

The SDFT algorithm may get considerable time reduction in comparison with DFT, FFT and 2D SDFT algorithms. The various CIF video sequences performance comparison is given in Table 3. For all the video sequences, processing time of the SDFT algorithm is 10.69% less than the 2D SDFT. The experimental results show that this SDFT algorithm always gives outperforming results than the other algorithms.

We have determined the processing time by giving the 4CIF video as input, and the experimental results are shown in Table 4. This shows the SDFT method outperforms by 37.30% in comparison with the 2D SDFT algorithm.

We have also determined the processing time by giving the QCIF video as input and the experimental results are shown in Table 5.

4 Conclusion

After comparing discrete Fourier transform and fast Fourier transform algorithms, it is found that this method gives better results in higher resolution videos. The 2D sliding DFT method cannot be functional on the top or left blocks of image because of the requirement to access the foregoing pixel arrangement. As we increase the resolution of the image ratio of blocks on image borders will reduce. We can expect that this algorithm will perform better for high-resolution videos.

References

1. Farhang-Boroujeny B, Gazor S (1994) Generalized sliding FFT and its application to implementation of block LMS adaptive filters. IEEE Trans Sig Process 42(3):532–538
2. Park C-S (2015) 2D Discrete Fourier transform on sliding windows. IEEE Trans Image Process 24(3):901–907
3. Jacobsen E, Lyons R (2003) The sliding DFT. IEEE Sig Process Mag 20(2):74–80
4. Duda K (2010) Accurate, guaranteed stable, sliding discrete Fourier transform [DSP tips and tricks]. IEEE Sig Process Mag 27(6):124–127
5. Farhang-Boroujeny HB, Lim V (1992) A comment on the computational complexity of sliding FFT. IEEE Trans Circ Syst-II Analog Digital Sig Process 39(12):875–876
6. Rosendo Maias JA, Exposito AG (1998) Efficient moving-window DFT algorithms. IEEE Trans Circ Syst -II: Analog Digital Sig Process 45(2):256–260
7. Deng G, Ling A (1996) A running Walsh-Hadamard transform algorithm and its application to isotropic quadratic filter implementation. In: Proceedings of European signal processing conference (EUSIPCO)
8. Mozafari B, Savoji MH (2007) An efficient recursive algorithm and an explicit formula for calculating update vectors of running Walsh Hadamard transform. In: Proceedings of IEEE 9th international symposium on signal processing and its applications (ISSPA)
9. Ben-Artzi G, Hel-Or H, Hel-Or Y (2007) The gray-code filter kernels. IEEE Trans Pattern Anal Mach Intell 29(3):382393
10. Ouyang W, Cham W-K (2010) Fast algorithm for Walsh Hadamard transform on sliding windows. IEEE Trans Pattern Anal Mach Intell 32(1):165–171
11. Jacobsen E, Lyons R (2004) An update to the sliding DFT. IEEE Signal Process Mag 21(1):110111
12. Park C, Ko S (2014) The hopping discrete Fourier transform [sp tips and tricks]. IEEE Sig Process Mag 31(2):135–139
13. Park C-S (2014) Recursive algorithm for sliding Walsh Hadamard transform. IEEE Trans Sig Process 62(11):28272836
14. Banks K (2002) The Goertzel algorithm. Embed Syst Program Mag 15(9):3442

Automated Lung Nodules and Ground Glass Opacity Nodules Detection and Classification from Computed Tomography Images

Vijayalaxmi Mekali and H. A. Girijamma

Abstract Lung cancer health care community depends on lung cancer Computer Aided Detection system to draw useful lung cancer details from Computed Tomography lung images. Nodules growth rate indicates the severity of the disease, which can be periodically radiologist analyzed by nodule segmentation and classification. Main challenges in analyzing nodules growth rate are lung nodules of different type requires special methods for segmentation, their irregular shape, and boundary. In this paper, automatic three-phase framework for lung nodules and nodules of ground glass opacity detection followed by classification is proposed. In this work, nodule segmentation framework uses proposed automatic region growing algorithm that selects set of black pixels as seed points automatically from output binary image for lung parenchyma segmentation followed by artifacts removal to reduce disease search space. Nodules are segmented based on nodule candidates center pixels identification and intensity feature of lung nodule candidates. Segmented nodules are classified using SVM classifier and classification results are compared with other considered classifiers KNN, boosting and decision tree. In the evaluation step, it was found that SVM classifier's performance is outstanding compared to other considered classifiers in this work. Complete automation in nodule detection within very less time is the key feature of the proposed method. CT images are taken from Lung Image Database Consortium and Image Database Resource Initiative (LIDC/IDRI) public database to evaluate the performance of proposed work. An accuracy of 98% (45/46) with less computational time is achieved. The experimental results demonstrated that the proposed method achieve efficient and accurate segmentation of lung nodules and ground glass opacity nodules with less computation time.

V. Mekali (✉)
Computer Science and Engineering Department, KSIT, Bangalore, India
e-mail: duruth.viju@gmail.com

H. A. Girijamma (✉)
Computer Science and Engineering Department, RNSIT, Bangalore, India
e-mail: girijakasal@gmail.com

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_127

1 Introduction

Lung cancer is carcinoma disease with high death rate compared to other cancers all over the world in human beings. Lung cancer killing rate is increasing day by day, approximately 1.6 millions of people globally per year. The main target for lung cancer is an elderly group of age 65 or older. Lung cancer detection study from various researchers showed that high curable rate is possible only by means of its early stage detection [1–3]. Thus, today also lung cancer detection research takes highest prominence in research and health care community. Two classes of lung carcinoma are Non-Small Cell Lung Cancer (NSCLC) and Small Cell Lung Cancer (SCLC). Most of the detected lung cancer (approximately 80%) is of type NSCLC and it is SCLC in less cases (approximately 20%). Assignment of stages to detect lung cancer is very important to determine its severity. TNM (Tumor, Nodule, Metastasis) staging is internationally accepted staging tool to stage different types of cancers [2].

Among all the medical image modalities, Low Dose Computed Tomography (LDCT) is an effective and proven GOLD STANDARD medical imaging modality that provides useful information about lung cancer by generating huge volume of image slices. With latest advanced technologies LDCT is more reliable and faster to detect very small nodules on lung area [4]. Lung nodules appear as high-intensity circular opacity with the smooth or irregular boundary of diameter 2–30 mm on LDCT [1]. Lung nodules are of different types of well-circumscribed nodules (with smooth boundary), irregular shape nodules, and nodules attached to other parts (juxta-vascular nodules attached to blood vessels and juxta-pleural nodules attached to lung pleura), solid, partly solid, nonsolid nodules, and Ground Glass Opacity (GGO). Presence of GGO on lung for long time indicates cancer. Figure 1 shows different types of nodules in CT images.

In one single scan, LDCT generates a huge amount of pathology's image slices of the area and those huge slices data set complicates radiologist work to perform manual analysis for detection of nodule pathology. Advanced Computer-Aided Detection/Diagnosis (CAD) systems based on various image processing, artificial intelligence, and machine learning algorithms are playing a prominent role in health care

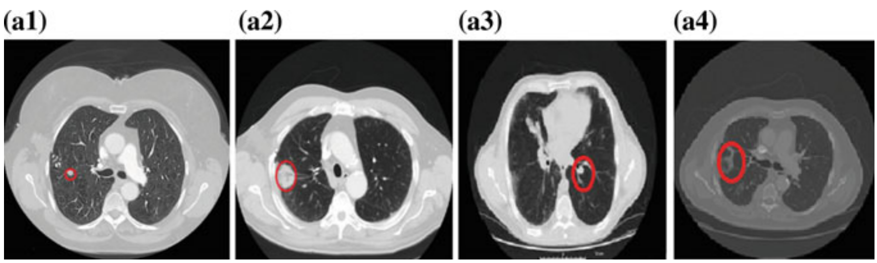


Fig. 1 Different types of lung nodules. **a**1 Circumscribed lung nodules. **a**2 Juxta-pleural nodule. **a**3 Juxta-vascular nodule. **a**4 GGO nodule

community to assist the radiologist in lung nodules detection of lung CT (Computed Tomography) images.

The major steps of lung cancer CAD system are lung parenchyma segmentation, lung nodule detection and classification of nodules as benign or malignant. Lung parenchyma segmentation is a difficult task. The main reasons are lung lobes shape is not same in all patients, high variation in geometric aspect of lung in different patients, existence of juxta-pleural and juxta-vascular nodules, overlap of chest wall, connected lung lobes. Lung region segmentation accuracy determines CAD system's nodule detection accuracy.

The rest of the paper is organized as follows.Section 1 is about lung cancer, LIDC dataset, usage of CAD system in lung cancer detection. Section 2 gives previous related work, Sect. 3 describes proposed methodology, and Sect. 4 discusses results and conclusion.

2 Related Work

Number of researchers has developed CAD systems for automatic detection of lung nodules using CT lung images and some are presented below.

Wang et al. [5] proposed lung nodules segmentation method based on 3D-extended Dynamic Programming (DP). In this work, non-nodule segmentation error rate was decreased by multidirection fusion techniques. First and second dataset set from LIDC were used and obtained true positive fraction of 75% and 71% for first and second dataset, respectively. Dehmeshki et al. [6] proposed method for automated lung nodules detection based on shape-based genetic algorithm. Spherical-oriented convolution-based filter was used in preprocessing step for input image enhancement. Seventy CT scan images were used to evaluate the proposed method, and method achieved about 90% nodule detection rate. In [7], author proposed CAD system based on fuzzy thresholding for parenchyma segmentation. For nodule segmentation and detection author used local Gaussian and mean curvatures based shape indexed map and local intensity dispersion information. Resultant CAD system was evaluated on 108 Clinical CT dataset and 90.2 detection rates were achieved.

Liu et al. [8] proposed CAD system was based on new algorithms for lung parenchyma segmentation and ROI extraction using circle shape descriptor from CT images. In this work, lung nodule candidates were predicted using ADE-Co-Forest algorithm. The resultant CAD system sensitivity was very high with low false positive. Novel CAD system for the detection of nodules attached to blood vessels from CT images was proposed by Tan et al. [9]. The author applied enhancement filters on nodule and vessel parts for nodule segmentation. To separate the attached blood vessels from nodules genetic algorithm and artificial neural network were used. Results are evaluated on CT images from LIDC.

In [10], Messay et al. the author presented a threshold and morphological operation based new fully automated CAD system for nodule detection. From each segmented nodules, 245 features were extracted to compare classifiers Fisher Linear

Discriminant classifier (FLD) and quadratic classifier. Classification results showed that FLD performance was better. Resultant CAD system used CT scans from LIDC and obtained 80.4% sensitivity. Adaptive thresholding based automatic Solitary Pulmonary Nodule (SPN) segmentation algorithm was proposed in [11]. In this work, extraction of nodule candidates method was achieved with compactness feature along with histogram analysis. True SPN from input CT images were classified using SVM classifier algorithm. In [12], the author presented fast lung nodules detection method based on thresholding for lung parenchyma segmentation, blood vessel filling by morphological operations, cylindrical shape filter for nodule enhancement, and SVM for false positive nodules reduction. Campos et al. [13] proposed novel supervised segmentation framework based on three features sets for detection of both solid and GGO lung nodules. The method used region growing algorithm for lung parenchyma segmentation and three features sets to segment lung nodules detection. The author used shape features for first type nodules segmentation, convergence matrix for second type nodules segmentation, and feature set region regression method for third type nodules segmentation. The output of three primary nodule segmentation is input for ANN to obtain accurate segmentation. For GGO nodules test, data set of CT images were taken from Nelson's trail database and CT images with solid nodules from hospital Sao Joao in Oporto, Portugal. Result was satisfactory with Jaccard coefficient 83% and dice coefficient of 93%. In [14], the author presented used region growing algorithm and Hassian algorithm for multisegmentation of lung region and lungs nodule connected vascular tree segmentation, respectively. Combined classifiers were used to classify the extracted nodule and achieved 98% of classification accuracy. Beigelman-Aubry et al. [15] proposed a solid nodule detection CAD system from Multidetector CT (MDCT) scan lung images. Along with the detection of solid nodules, other key features of this CAD were evaluation of tracking and reading time.

In this work, nodule segmentation framework uses proposed automatic region growing algorithm that selects set of black pixels as seed points automatically from output binary image for lung parenchyma segmentation followed by artifacts removal to reduce disease search space. Nodules are segmented based on nodule candidates center pixels identification and intensity feature of lung nodule candidates. Segmented nodules are classified using SVM classifier and classification results are compared with other considered classifiers KNN, boosting and decision tree.

3 Materials and Methods

The proposed CAD system consists of stages such as (1) Lung lobes segmentation (2) Lung lobes boundary refine (3) Nodules detection followed by classification (4) Comparison of classifiers performance.

3.1 LIDC Data Set

CT lung images are taken from available public Lung Image Database Consortium-Image Database Resource Initiative (LIDC-IDRI). It is internationally accepted CT lung images database. National Cancer Institute (NCI) founded LIDC in 2004 in collaboration with multiple hospitals. LIDC contains lung CT slices of size 512 * 512 with thickness 1.25–2.5 mm and pixel size in range 0.48–0.72 mm. All LIDC CT images are in Digital Imaging and Communications in Medicine (DICOM) format. The aim of this database is to promote researchers from all over the world to develop and evaluate CAD system for lung nodule detection [4, 16].

3.2 Lung Nodules Detection Methodology

Proposed methodology consist of lung parenchyma segmentation and separation of lung lobes if they are attached, the next step is artifacts removal such as mediastinum, trachea, and CT examination bed, third step is to detect lung nodules, and the last step is a classification of nodules as benign or malignant. Entire CAD system accuracy depends on nodule classification accuracy which in turn depends on the accuracy of lung segmentation. Figure 2 shows the block diagram of the proposed work.

3.3 Lung Segmentation

Objects in the input CT image are separated from image background by applying iterative thresholding algorithm given in Algorithm 1. The output generated is the binary image with objects and background information. Figure 3 explains the original input CT image, histogram of input image, and output of iterative thresholding. Binary image contains artifacts such as thorax region, CT examining bed.

Algorithm 1: Iterative Thresholding algorithm

Input: CT lung image
Output: Threshold image
Step 1: Let T_h is highest grey level in input image T_m is lowest gray level in input image T_h and T_m are obtained from histogram of input image.

$$T=(T_h+ T_m) / 2 \quad (1)$$

Step 2: Segment the image using T

$CT(x,y) > T$	0	Foreground	F_g
$CT(x,y) < T$	1	Background	B_g

Step3: $T_i=(F_g+B_g)/2 \quad (2)$

Step 4: Repeat steps 2 and 3 until Ti-Ti+1 small enough that further thresholding generates output with no further changes.

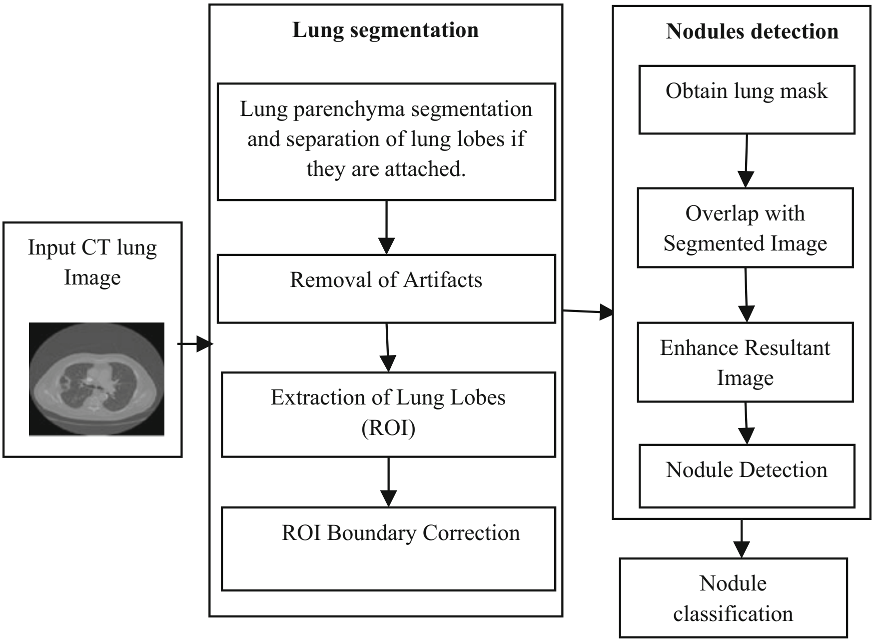


Fig. 2 Proposed CAD system to detect and classify lung nodules

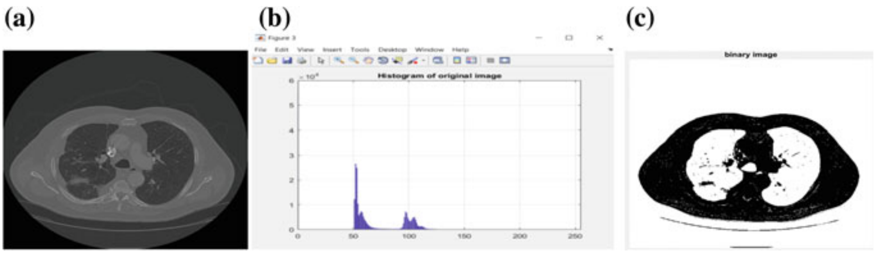


Fig. 3 **a** Original input CT lung image. **b** Histogram of input image. c Output binary image

Output of Algorithm 1 contains artifacts. To reduce search space for nodule detection these artifacts are cleared by first clearing the boundary of binary image followed by automated seed selection region growing algorithm given in Algorithm 2. Algorithm 2 does not require user interaction to select seed points and it automatically calculates seed points to output possible regions in binary input image. Border and artifacts cleared binary lung output image is shown in Fig. 4.

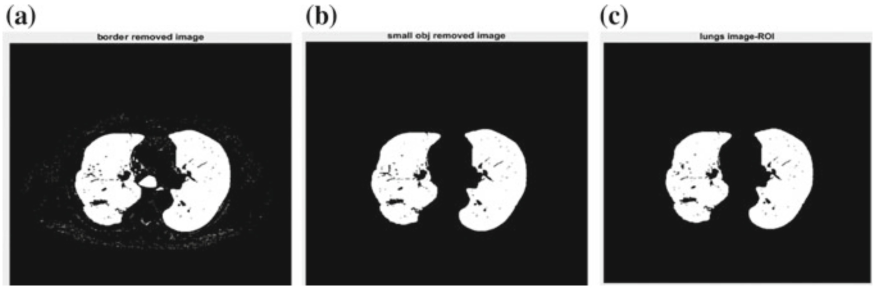


Fig. 4 **a** Binary image without border. **b** Output of automatic region growing method without artifacts and with removed small objects. **c** Extracted lung ROI

Algorithm 2: Automatic seed selection region growing algorithm

Input: Output Threhold image of Algorithm1
Output: Segmented image

Step 1: Sequences of black color pixels are identified by scanning binary input image in horizontal direction starting with leftmost pixel to right most pixel and vertical direction starting top most pixel to bottom most pixel. First black pixel in the set is considered as seed point for region growing.

Step 2: Let UA set of is all unallocated pixels and R_i possible seed regions,

$$UA = \{x \notin \bigcup_{i=1}^{n} Ri : P(x) \cap \bigcup_{i=1}^{n} Ai \neq \emptyset\}, \quad (3)$$

Where I = (1,2,……..n) and P(x) is either 4 connected or 8 connected neighbors of pixel x

Step 3: If P(x) and R_j intersect, then calculate δ(x) ie similarity or difference measure between x and Intersection region R_j and δ(x)is defined as

$$\delta(x) = |gv(x) - mean\{gv(z)\}| \quad (4)$$

where gray level at pixel x is gv(x) and mean gray level of all pixels z in the region R_j is gv(z). and j=(1,2,3…….n). Eq. 4 is calculated for all the intersected regions R_j with P(x) and minimum δ(x)is considered for inclusion of x to that region.

Step 4: Repeat Step 3 for all x ϵ UA, add a pixel y that satisfies the Eq. 3

Lung Nodules Segmentation

Lung nodule detection process consists of two steps first identify the center pixel of nodule candidates and second detect the nodules based on intensity. In segmented image, lung lobes and nodule candidates are with black edges. Noncancerous (<3 mm, benign) nodules are smaller in size compared to cancerous nodules (>3 mm, malignant). Thus, nodule size calculation is also a key feature of our nodules detection method. Black pixels collection in Fig. 4 shows the possibility of nodule candidates. Algorithm 3 first determines the center pixel of nodule candidates. In the second step using center pixels approximate elliptical boundary for each nodule candidate is identified to calculate the average intensity of pixels within the boundary. As in CT lung image, nodules appears as white spot with higher intensity value. Algorithm 3 uses this intensity feature to locate the exact nodules in the segmented image.

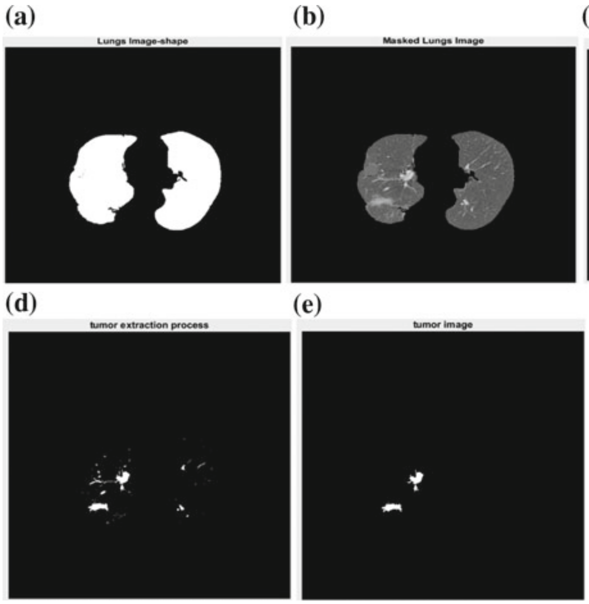


Fig. 5 **a** Lungs shape of input image. **b** Masked lung image. **c** Enhanced image. **d** Lung nodule candidates image. **e** Image with extracted nodules

Nodule candidates with average pixel intensity greater than a specified threshold are extracted and are correctly identified in lung nodules. Figure 5a–d shows nodules segmentation process of our method, Fig. 5e is output image with extracted nodules.

Algorithm 3: Nodule Segmentation algorithm

Input: Edge detected image
Output: Image with segmented nodules
Step 1: Begin with black edge pixels,
    for each considered black edge pixels
        Determine 8 – connected neighborhood black pixels and add to center pixels
    end /*Results in image CN */

Step 2: Obtain the mask image M of original input image.
Step 3: Superimpose CN with image M to obtain superimposed image S.
Step 4: Enhance the resultant image S of Step 3 to obtain enhanced image E.
Step 5: For nodule candidate i with center pixels Px_i in image E
        a) Obtain elliptical boundary region.
        b) Calculate the average of pixels intensity AP within the nodule candidate.
            if AP> NT, where NT is specified intensity to segment the nodules
                nodule candidate is correctly identified as lung nodule
                and set its center pixels as Cx_j.
            end if
    end for

3.4 Nodules Classification

The last stage of our work is a binary classification of extracted nodules. To classify the nodules, two intensity-based features sets with totally 17 features are used. First set includes five histogram-based textural features: Mean, Variance, Entropy, Energy, and Skewness. The second set includes 12 features based on Gray Level Co-variance Matrix (GLCM). Gray level of input lung CT image is used to confirm the GLCM size. Based on pixels relationship in four direction with angular difference of 45°: 0°, 45°, 90°, 135°, and distance multiple GLCM offsets are calculated. Twelve GLCM-based textural feature set includes: Sum and difference mean correlation, Inertia, Cluster tendency and shade, Autocorrelation, Maximum Probability, Homogeneity, Standard deviation, Sum and difference entropy. These considered features are used as input to SVM and other comparing classifiers. Final output image, Fig. 6a shows lung with cancer disease and Fig. 6b gives extracted nodules condition.

3.5 Classifier Validation

K-fold validation framework along with Matthews's Correlation Coefficient (MCC) given in Eq. 5 is applied to validate binary classification of SVM-RBF kernel and other considered classifiers KNN, boosting, decision tree in our classification work. Table 1 describes parameters used for MCC calculation.

$$\text{MCC} = \frac{(Tp * Tn) - (Fp * Fn)}{\sqrt{(Tp + Fp) * (Tp + Fn) * (Tn + Fp) * (Tn + Fn)}} \quad (5)$$

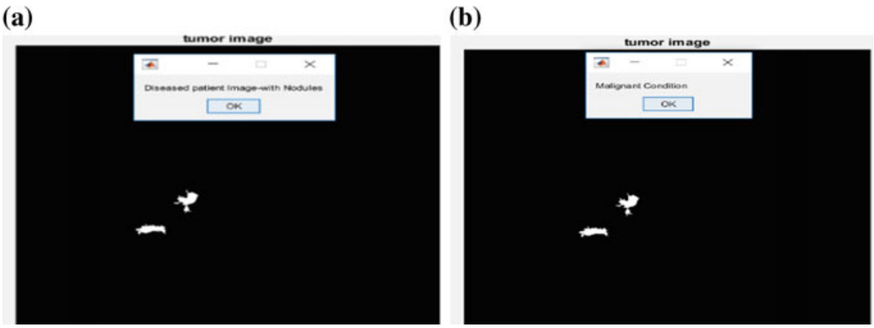


Fig. 6 **a** and **b** Output image with extracted nodules

Table 1 Values of *Tp*, *Tn*, *Fp*, and *Fn*

	Malignant	Benign
Malignant	*Tp*	*Tn*
Benign	*Fp*	*Fn*

where *Tp*, *Tn*, *Fp,* and *Fn* are given in Table 1.

4 Result

Input CT images are taken from LIDC datasets and Fig. 7 shows extracted lung nodules from different input lung CT slice. In Fig. 7, upper row is the input image and lower row shows nodules segmented from input images. Totally 17 features are used to perform binary classification (benign or malignant). Receiver Operating Characteristic (ROC) curve in Fig. 8 shows SVM classifier's outstanding performance compared to other considered classifiers. Table 2 provides the accuracy and AUC values obtained using classifiers SVM-RBF kernel, boosting, decision tree, and KNN.

5 Conclusion

In this work, nodule segmentation framework uses proposed automatic region growing algorithm that selects a set of black pixels as seed points automatically from output binary image for lung parenchyma segmentation followed by artifacts removal to

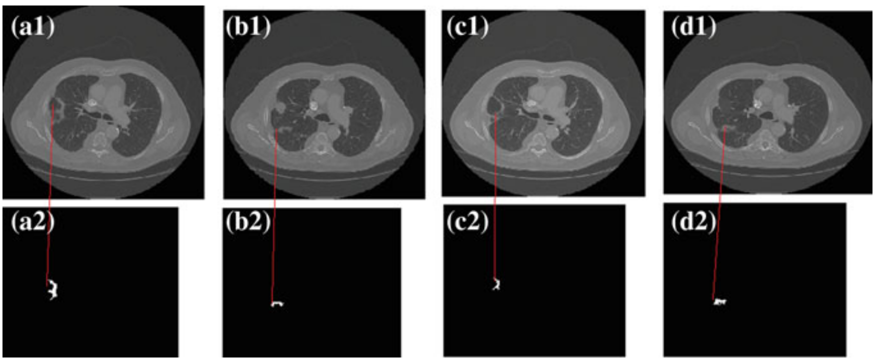


Fig. 7 **a**1 **b**1 are input lung CT images with lung nodules and **c**1 **d**1 with GGO nodules **a**2 **b**2 **c**2 **d**2 are segmented nodule images

Table 2 Comparison of classifiers performance in the proposed system

Sl. No	Classifiers	Accuracy (%)	AUC (%)
1.	SVM-RBK	98	96
2.	Boosting	90	88
3.	Decision tree	77.45	75
4.	KNN	74.13	71

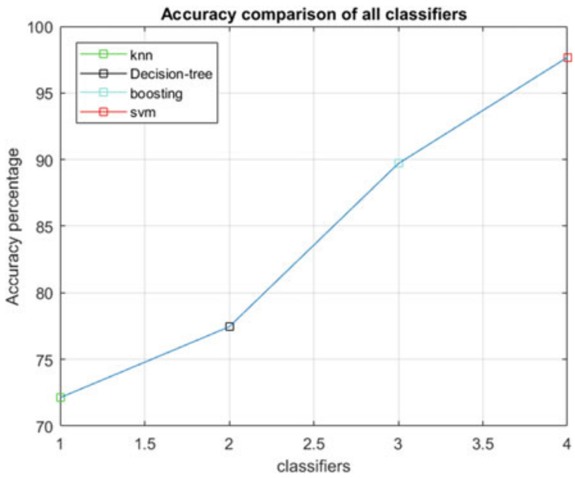


Fig. 8 Receiver operating characteristic curve of classifiers

reduce disease search space. Nodules are segmented based on nodule candidates center pixels identification and intensity feature of lung nodule candidates. Segmented nodules are classified using SVM classifier and classification results are compared with other considered classifiers KNN, boosting and decision tree. In the evaluation step, it was found that SVM classifier's performance is outstanding compared to other considered classifiers in this work. Complete automation in nodule detection within very less time is the key feature of the proposed method.

Acknowledgements I thank my research guide Dr. Girijamma H A, Professor, Department of CSE, RNSIT, Bangalore, India. For supporting to complete this research article. I would like to thank public LIDC database from which images are taken to carry a reach work.

References

1. Non communicable Diseases Progress Monitor (2017) Geneva: World Health Organization; 2017. Licence: CC BY-NC-SA 3.0 IGO
2. Wender R, Fontham ETH, Barrera E (2013) American cancer society lung cancer screening guidelines, National Institute of Health, CA Cancer J Clin 63(2):107–117. https://doi.org/10.3322/caac.21172, (2013)
3. Hillary Wasserman Becky Bunn.: Lung Cancer Facts and Statistics, International Association for the Study of Lung Cancer(IACLS), WCLC, 20173999999 (2017)
4. Habib MSL (2009) A computer aided diagnosis system (CAD) for the detection of pulmonary nodules on CT scans. Systems and biomedical engineering Department, Faculty of Engineering, Cairo University, Giza, Egypt
5. Wang Q, Song E, Jin R, Han P, Wang X, Zhou Y, Zeng J (2009) Segmentation of lung nodules in computed tomography images using dynamic programming and multidirection fusion techniques. Acad Radiol 16(6):678–688
6. Ye X, Lin X, Dehmeshki J, Slabaugh G (2009) Shape-based computer aided detection of lung nodules in thoracic CT images. IEEE Trans Biomed Eng 56(7)

7. Ye J, Lin XY, Dehmeshki J, Slabaugh G, Beddoe G (2009) Shape based computer aided detection of lung nodules in thoracic CT images. IEEE Trans Biomed Eng 56:1810–1820. https://doi.org/10.1109/tbme.2009.2017027
8. Liu Y, Xing Z, Deng C, Li P, Guo M (2010) Automatically detecting lung nodules based on shape descriptor and semi-supervised learning. In: International conference on computer application and system modeling. IEEE, https://doi.org/10.1109/iccasm.20https://doi.org/10.5619447, Taiyuan, China
9. Tan M, Deklerck R, Bister BJM, Cornelis J (2011) A novel computer-aided lung nodule detection system for CT images. Med Phys 38(10)
10. Messay T, Hardie RC, Rogers SK (2010) A new computationally efficient CAD system for pulmonary nodule detection in CT imagery. Med Image Anal 14(3):390–406. https://doi.org/10.1016/j.media
11. Shao H, Cao L, Liu Y (2012) A detection approach for solitary pulmonary nodules based on CT images. In: IEEE, 2nd international conference on computer science and network technology (ICCSNT). Changchun, pp 1253–1257
12. Teramoto A, Fujita H (2013) Fast Llung nodule detection in chest CT images using cylindrical nodule-enhancement filter. Int J Comput Assist Radiol Surg 8(2):193–205. https://doi.org/10.1007/s11548-012-0767-5
13. Campos DM, Simoes A, Ramos I, Campilho A (2014) Feature-based supervised lung nodule segmentation. In: IFMBE proceedings the international conference on health informatics 42, 23 https://doi.org/10.1007/978-3-319-03005-0_7. Springer International Publishing, Switzerland
14. Elsayed O, Mahar K, Kholief M, Khater HA (2015) Automatic detection of the pulmonary nodules from CT images. IEEE. doi: https://doi.org/10.1109/intellisys.2015.7361223. SAI Intelligent Systems Conference (IntelliSys)
15. Beigelman Aubry C, Raffy P, Yang Ronald W, Castellino Philippe A, Grenier A (2007) Computer-aided detection of solid lung nodules on follow-up MDCT screening: evaluation of detection, tracking, and reading time. Chest Imaging AJR:189 https://doi.org/10.2214/ajr.07.2302
16. https://wiki.cancerimagingarchive.net/display/Public/LIDC-IDRI

A Review—Edge Detection Techniques in Dental Images

Aayushi Agrawal and Rosepreet Kaur Bhogal

Abstract Edge detection plays an important role in digital image processing applications. The main aim of edge detection is to identify the discontinuity in images, where the sharp changes in intensity take place. This research work presents the edge detection technique in dental X-ray images (panoramic radiograms), which is advantageous to separate teeth individually for better classification and identification of diseases. The objective is to study and compare the various algorithms that are Sobel, Prewitt, Canny, multiple morphological gradient (mMG), line analyzer, neural network, genetic algorithm, and infinite symmetric filter (ISF), multi-scale and multi-directional analysis with statistical thresholding (MMST), and fuzzy logic approach for edge detection in dental X-ray images. There are many difficulties in finding diseases from panoramic dental images only, and hence to overcome these difficulties edge detection is introduced. Some of the dental diseases that require edge detection for their identification are discussed. Based on capability of detecting the diseases accurately and total number of diseases detected from the dental images by the use of edge detection, comparison of results takes place.

1 Introduction

Edges can be specified as a group of immediate pixel location where a sudden variation in intensity values takes place. Edges are the boundaries among objects and background. An edge detector can be used for feature extraction, image segmentation, and object identifications [1].

In dental X-ray imaging, edge detection technology plays a very helpful role in detecting and diagnosis of the diseases. Dental radiograms are poor and complicate in some of the diseases extraction such as tooth decay, cavities [2], tooth abscess,

A. Agrawal (✉) · R. K. Bhogal
Electronics and Communication Engineering, Lovely Professional University, Phagwara, India
e-mail: aayushiagrawal20495@gmail.com

R. K. Bhogal
e-mail: rosepreetkaur12@gmail.com

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_128

impact tooth, etc. Brightness of X-ray image is not good enough because of sequins occurring due to the presence of water on teeth.

Neural network method is used for detection of edges in dental X-ray images where inputs to the network are pixels of original image after minimizing the output error, and output pixels are replaced in the edge-detected image. Artificial neural network (ANN) is used because in this iterative training and learning takes place from input to output mapping. This algorithm is helpful in finding problems such as tooth decay, bone damage used for supporting the teeth, and dental injuries. This method may be extended for the diagnosis of major problems related to teeth [3].

Firstly, X-ray image required preprocessing for enhancement using Gaussian low-pass filter, then various edge detection techniques are applied and at last comparison of results takes place, which shows that Canny operator provides best results among Sobel and Prewitt [4]. After preprocessing of image, better edge detection techniques might be applied for detection of true edges in dental images. The algorithm of multiple morphological gradient (mMG) is applied in [5] for clear visibility of boundaries of object in panoramic radiograms. Some positions of dental caries are also defined, which says that it is divided into two groups, that are smooth surface and gap of teeth. The result of this algorithm is very useful for the identification of cavities and verification of the same is done by two dentists. Furthermore, this algorithm should be applied for the detection of more number of dental diseases.

Before edge detection in dental X-ray images the two steps that are to be followed are image enhancement and teeth segmentation so that the proper identification of every teeth occurs [6]. The problem of finding missing teeth [7] is to be solved using this procedure of edge detection. In jaw, where the teeth is missing, values of standard deviation, Euler number, and area become zero. Additionally, for proper alignment of teeth, its measurement is very important. Traditionally, gear tooth micrometer and vernier caliper are used but results are not in fraction, then minute changes in measurement lead to improper alignment of teeth. Solution of this problem is given in [8] by the use of edge and gray-based method. The results prove that it is possible to place tooth correctly at exact position. Results might be improved by soft algorithm edge detection techniques like genetic algorithms (GA) that provide flexible edge detection in very shady images [9]. Since it is also applied in dental X-ray images for identifying the diseases like tooth decay, impact tooth, and so on. Mainly GA consists of three stages; they are selection, crossover, and mutation for processing of any application. Comparison of results takes place with Roberts [10] operator and found that GA gives better results comparatively.

There is less amount of radiation in cone beam computed tomography (CBCT) as compared to panoramic radiograms because this beam is divergent and forms a cone-like structure. For the edge detection of CBCT segment, standard neural network is used, where input to this network is some parameters of image that are used and one hidden layer is also present in between input and output layer. Results are compared with Canny edge detector and it proves that NN is fast comparatively and is helpful for providing 3D reconstruction of teeth [11].

Identification of dental abscess is also major problem by the use of dental X-ray images only. For this optimal edge detection required [12] solves the same problem which states that firstly image requires removal of noise from it and after that edge detection is applied for extraction of abscess. Results of this edge-detected images are shown to the dentist and well appreciated by them. And, for root canal treatment, finding the length of root is very important parameter and that is done by the use of edge detection in dental X-ray images. The implementation of algorithm for the solution of abovementioned problem takes place by the use of MMST where Laplacian pyramid is used to decompose the image into number of levels, directional decomposition of image is done by the use directional filter banks. Statistical thresholding is done for detection of edges in an image. And, results conclude the clear and precise detection of root canal length for further diagnosis [13]. Root length might be measured more accurately by soft algorithms where output is dependent on applied rules and membership function.

The edge detection by ISEF [14, 15] is helpful in detecting dental caries in tooth decay from dental X-ray images. After that it also gives the decision for the treatment like root canal treatment or filling. Results are giving the idea where the caries-affected area is present and also verified by doctor. Since 3D reconstruction of dental images plays an important role in identification of various diseases. Some of the challenging issues are solved which are found in CBCT segments that are connectivity, noisiness, position, topological changes, and low resolution by 3D dental images [16, 17] which is obtained by doing edge detection by the use of Canny method. For the validation of method, triangulation of the sparse points is used. Moreover, this 3D reconstructed image will be used for detection of major issues related to teeth.

Sobel and Canny edge detector [18] are widely used for detection of edges as they are simply calculating the gradients of image for edge detection. Here, it is useful in detecting edges for dental age assessment. Comparison of Sobel and Canny edge detector occurs and results that Sobel is better than Canny for dental images, and then soft algorithms are used where inputs are dealing with approximation model. [19–21] shows one of them, that is, fuzzy logic approach for the detection of edges from dental X-ray segment. Fuzzy inference system (FIS) [22, 23] is used which is having three main steps that are fuzzification, knowledge base, and defuzzification. Extraction of edges depends upon the degree of whiteness and blackness of eight neighbor gray level pixels that are done by powerful rules applied to it. The results are compared with Roberts method and it is clearly visible that more precise edges are detected by fuzzy logic approach. Higher level of fuzzy logic approaches might be used for better detection of edges.

The section of this paper is prepared as follows: the basic human teeth structure is demonstrated in Sect. 2. The common dental diseases along with causes, treatment, and prevention are described in Sect. 3. Overview of edge detection in X-ray is given in Sects. 4 and 5 comprises flowcharts of each technique reviewed in this paper. In Sect. 6, results and discursion take place on the basis of outcome produced by edge detection techniques. Finally, in Sect. 7 some conclusion and future scope according to the results takes place.

2 Human Teeth Structure

Human teeth structure shown in [23] Fig. 1 mainly consists of three tissues, they are as follows:

(1) *Hard tissues*: It is also known as calcified tissues having a hard intercellular matrix and also highly mineralized. Cementum, dentine, and enamel are the hard tissues in human teeth structure.
(2) *Soft tissues*: It is present to protect and cover the root of the teeth. Tooth pulp is soft tissue found in teeth.
(3) *Supporting tissues*: It is present around the teeth and is helpful in providing the necessary support to hold teeth in use. There are there supporting tissues and they are alveolar bone, gingiva, and periodontal ligaments.

3 Dental Diseases

The major significant function of teeth is to increase appearance and individual's integration in modern society. It is also helpful to make ready the food digestive. Sometimes when humans are not aware of the importance of teeth and not brushing their teeth properly and many more causes may lead to some dental diseases [24] which are mentioned in Table 1 along with its prevention and treatments.

These dental issues also have common risk causes as noncommunicable diseases. According to the importance of dental health, WHO organized world health day in 1995 year with the name "Oral health for healthy life" [25]. A survey of oral health

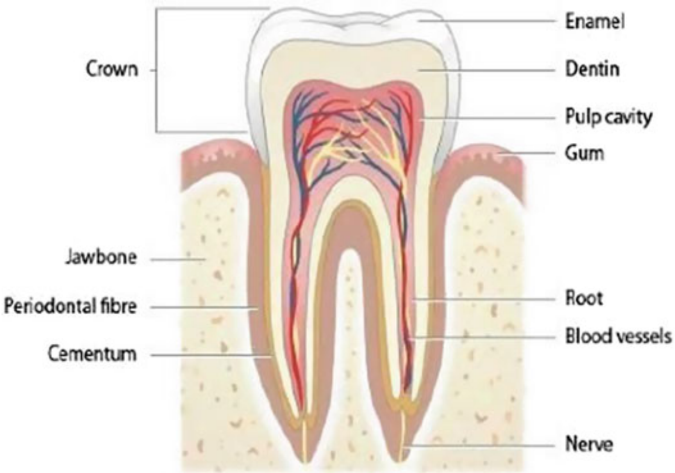
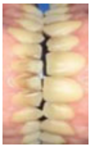


Fig. 1 Basic structure of human teeth

Table 1 Common dental diseases along its causes, prevention, possible treatment outcome, and images

S. No	Dental diseases	Causes	Prevention	Possible treatment outcome	Images
1.	Tooth decay	Improper teeth cleaning, over eating, and drinking of sugar	Washing mouth after eating, less consumption of sugar	Remineralization, fluoride treatments	
2.	Chipped tooth	Due to some accidents, cavities which make weaken the tooth	Uses of mouth guard, avoid eating the hardest food	Tooth filling, covered with crown (made up ceramics)	
3.	Stained teeth (Dis-coloration)	Chewing a tobacco or smoking, improper teeth brushing	Avoid chewing tobacco or smoking brush teeth twice in a day	Tooth bonding, advance whitening	
4.	Impacted teeth	Improper eruption of tooth, insufficient space in jaw	Continuously updating the dental X-ray	Tooth extraction	
5.	Sensitivity (Cold)	Due to some predefined problems and they are cavities, cracked, exposed	Avoid drinking or eating acidic food or drinks	Apply protection layer (fluoride layer), brush with fluoride gel	
6.	Hyperdontia	Actual cause is unknown but many experts believe that it is due to genetic	Not available	Surgical extraction, orthodontic treatment	

(continued)

Table 1 (continued)

S. No	Dental diseases	Causes	Prevention	Possible treatment outcome	Images
7.	Crooked teeth	Sucking of thumb, usage of baby bottle at large extent, teeth grinding	Avoid thumb sucking, avoid usage of baby bottle a lot	Mouth guard, therapy	
8.	Diastema	Different shape and size of teeth, tongue thrusting	Placement of permanent retainer	Frenectomy, dental crowns, invisalign	
9.	Periodontal diseases	Smoking, diabetes, hormonal changes in women	Avoid smoking, repeated checkup and cleaning of teeth	Flap surgery, bone and tissues grafting, deep cleaning	
10.	Bruxism	Stress, sleep disorders	Usage of bite splint at night	Biofeedback exercises, physical therapy, plastic tooth guard, relaxation techniques	
11.	Dental abscess	Smoking, diabetes, gum disorders	Brushing of teeth after every meals, avoidance of smoking	Root canal treatment, drainage of pus	
12.	Dental trauma	Hit by hard surface, rigorous sports activities	Usage of mouth guard	Tooth filling	

(continued)

Table 1 (continued)

S. No	Dental diseases	Causes	Prevention	Possible treatment outcome	Images
13.	Oral cancer	Smoking, tobacco	Avoidance of smoking and tobacco	Radiation therapy	

[26] demonstrates that it is important for every stage of life, mainly for older, adults, and seniors because of it various preventions that are needed to be followed which are discussed in next section of this paper.

4 Edge Detection in Dental X-Ray

Diagnosis of dental diseases is conventionally contained by radiographic films. Nowadays, dental X-rays came into existence for detection of several diseases such as impacted tooth, tooth decay, crooked teeth, abscess, and chipped tooth. But, the main issues with dental X-ray is that it is not very prominent. After this, edge detection techniques should be applied for better classification of dental diseases form X-ray image. Although edge detection carried out different methods as mentioned above but its main aim is to detect true edges for the proper identification of dental diseases. Several edge detection methods are discussed in the form of flowcharts in next section.

5 Experimental Analysis

Following are the methods mentioned in the form of flowcharts for detection of edges in dental X-ray images, which are discussed in Sect. 1.

Figure 2 shows line analyzer techniques for teeth, in this algorithm, image smoothing is applied for removal of noise by the use of Gaussian filter, gradients are calculated from Sobel operator. For converting blurred edge to sharp edge pixels non-maximum suppression is used, and finally double thresholding takes place to differentiate between weaker and stronger edges.

Figure 3 shows mMG method for edge detection, preprocessing of image that is cropping to extract the affected area only. Morphological operations are applied that are dilation and erosion. Dilated image is subtracted from eroded image to calculate the morphological gradient of the cropped image.

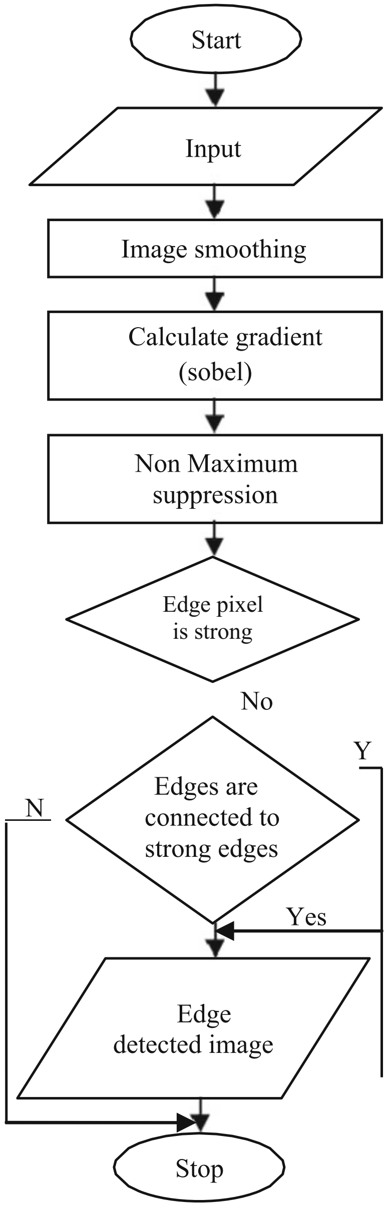


Fig. 2 Line analyzer technique for teeth

Figure 4 shows ANN training for edge detection, there were mainly two steps for detection that are pixel classification, which says value of pixels is a part of segment or not. Edge detection is done by the use of convolved mask obtained by the output of neural network.

Fig. 3 mMG method

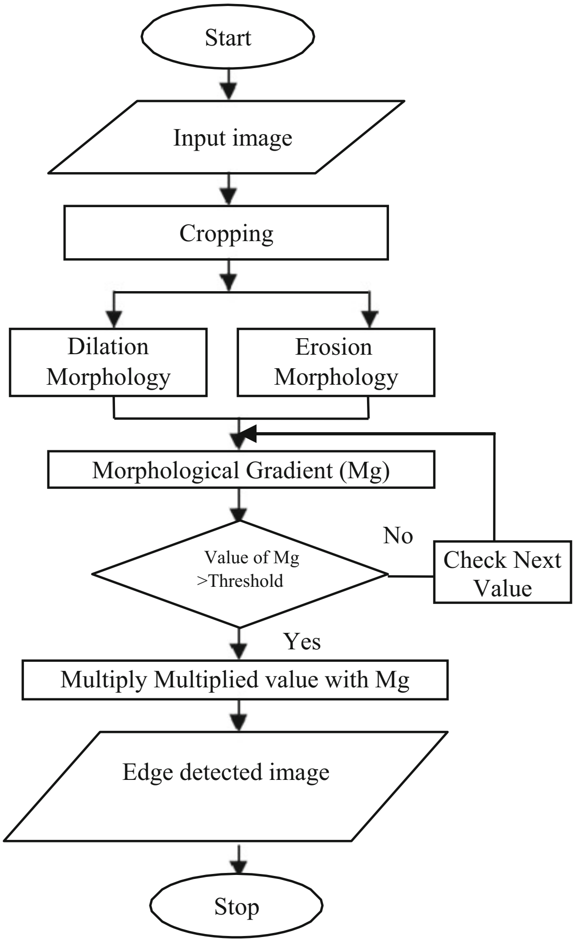
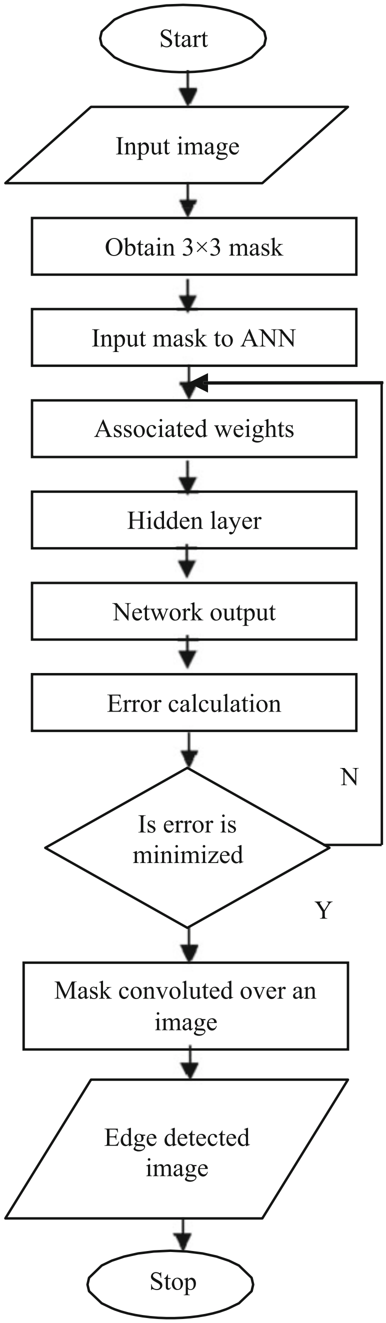


Figure 5 shows neural network and Canny algorithm for edge detection, in this both algorithms (Canny and neural network) combine to detect very accurate edge-detected image, and here output of Canny is given to input of neural network if edges are not properly detected.

Figure 6 shows ISEF method for abscess detection in dental images, and here filtering is done by the use of median filter for better enhancement of gray-scaled input image. ISEF filter is applied in both “x” and “y” direction. Binary Laplacian technique is applied by subtracting filtered image from original image. For detection for strong edges, hysteresis thresholding is used.

Figure 7 shows the method for 3D reconstruction of input image, and in this algorithm edges are detected using Canny method. Contours are detected in all direction to combine that section to form a 3D reconstructed image.

Fig. 4 Artificial neural network method



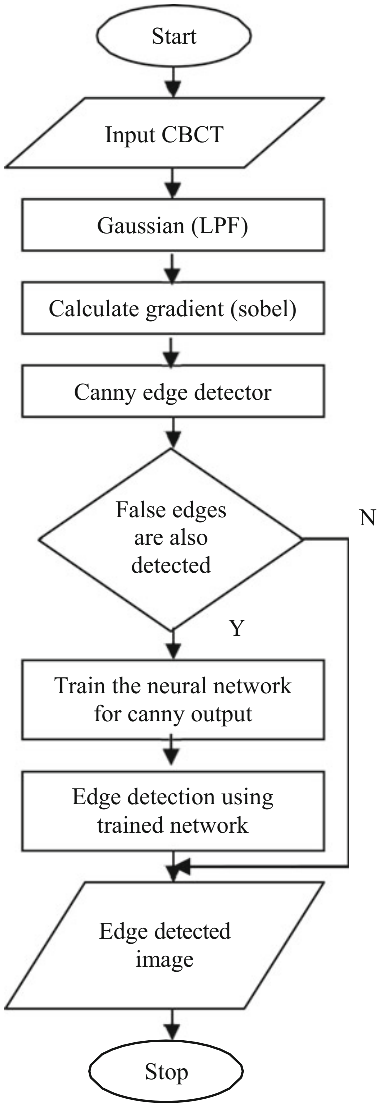


Fig. 5 Neural network method and Canny operator

Figure 8 shows the edge detection algorithm for endodontic working length measurement, and here edge detection takes place using MMST method. In that removal of speckle noise could be done by frost filter, adaptive equalization for enhancement takes place, and Laplacian pyramid decomposition is applied to decompose the X-ray image of tooth into number of levels.

Figure 9 shows the method for detection of desired teeth, and in this histogram equalization has occurred for enhancement of image. Otsu's method is used for

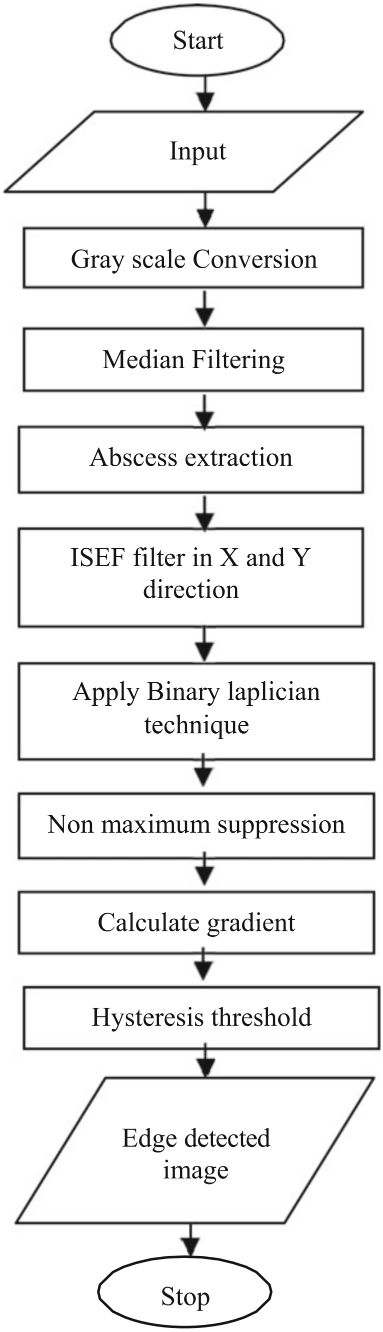
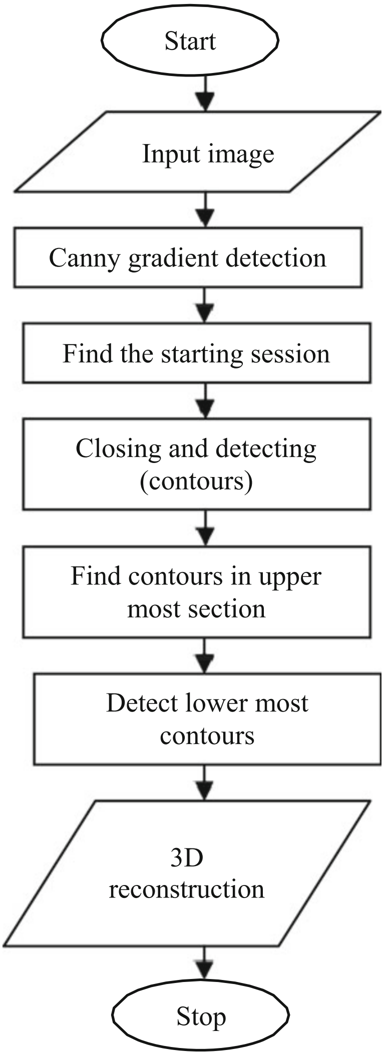


Fig. 6 ISEF method

Fig. 7 Canny edge detector

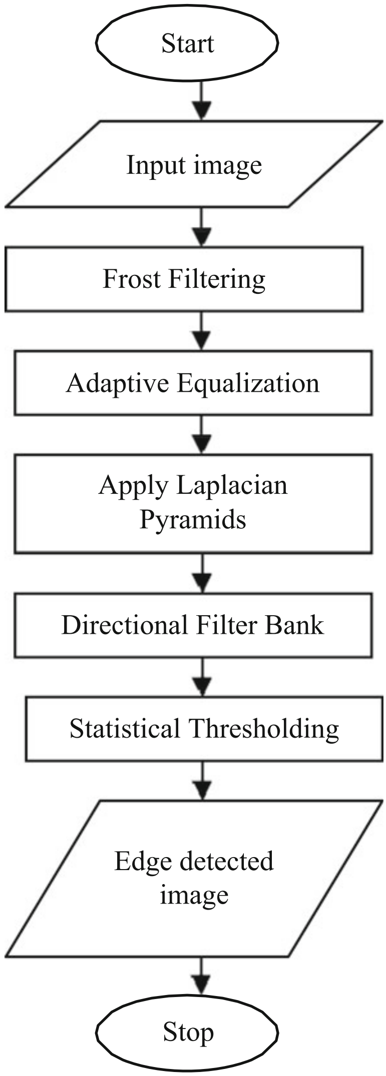


segmentation of desired area. Canny edge detection is used for detection of desired teeth.

Figure 10 shows preprocessing steps before edge detection, and here preprocessing is required for enhancement of input image, for that image resizing, grayscale conversion, and filtering by the use of Gaussian filter are done. Edges are more accurately found using Sobel or Prewitt or Canny edge detector.

Figure 11 shows fuzzy logic approach for edge detection, and here rules are dependable on eight neighbor gray level pixels, decision depends upon degree of blackness and whiteness, which decides that edges are there in the image or not.

Fig. 8 MMST method



Membership function is used where predefined rules are applied for accurate detection of edges from the image.

Figure 12 shows ISEF method for caries region detection, the algorithm shows the conversion of image into gray scale for further processing. Morphological operation and filtering are required for enhancement and still the caries-affected tooth is not found, then ISEF method for edge detection is applied.

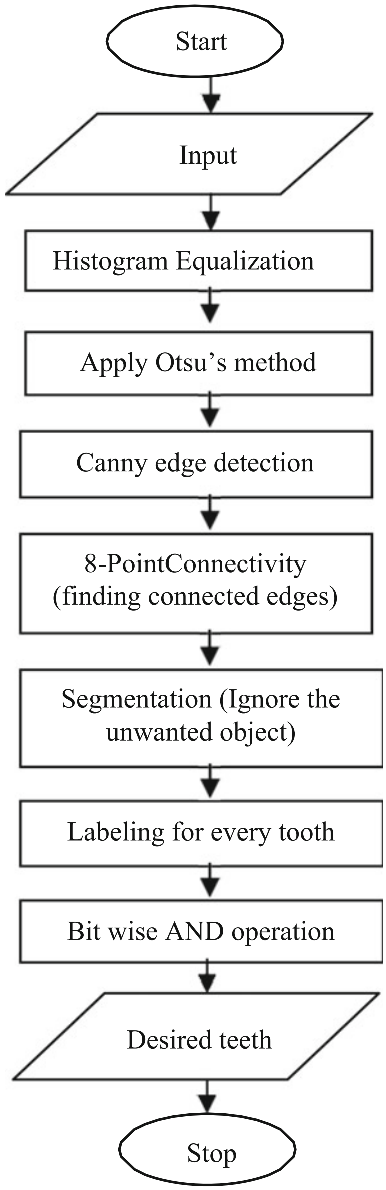
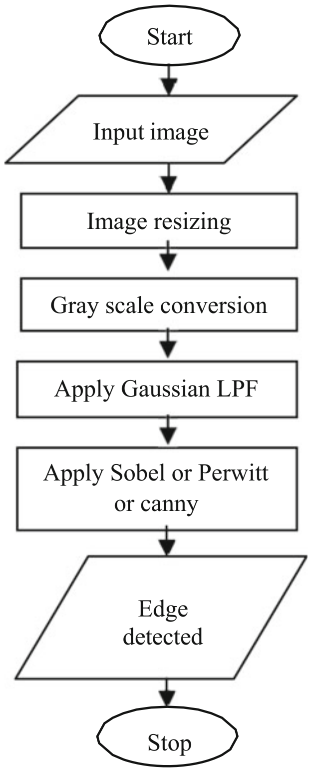


Fig. 9 Segmentation and Canny edge detector

6 Results and Discussion

The results of various techniques are compared, on the basis of ability to identify the type of dental diseases or solution of issues created in dental X-ray images

Fig. 10 Sobel and Prewitt



(bitewing X-ray, periapical X-ray, panoramic X-ray, and CBCT) [27] by the use of edge-detected dental images.

Table 2 shows the results of different edge detection techniques for particular dental diseases. Input to the algorithm is different type of X-ray images on the basis of total number of diseases detected, and results demonstrate that genetic algorithm, neural network, and fuzzy logic approaches give better results among all edge detection technique such as Sobel, Prewitt, Canny, mMG, and ISEF.

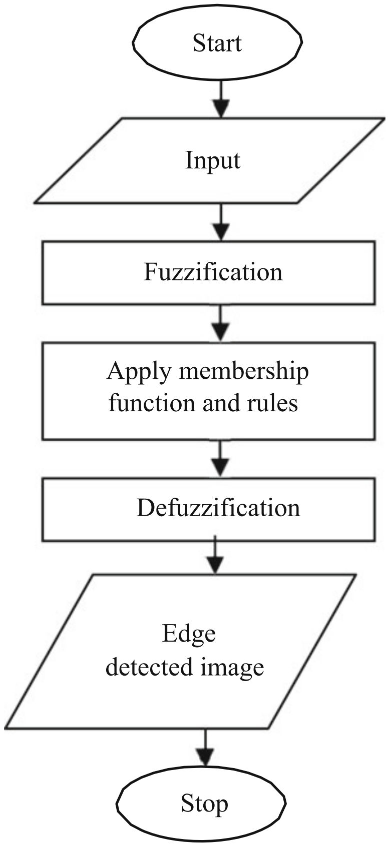


Fig. 11 Fuzzy method

7 Conclusion

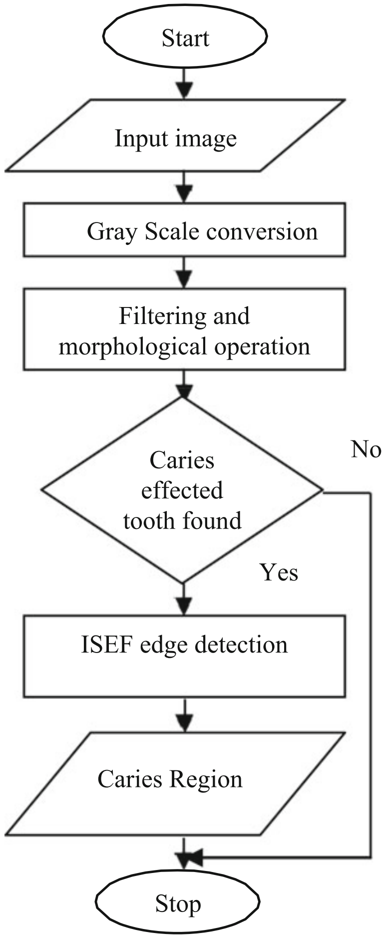
This paper is the review on several edge detection techniques applied on dental images. Each edge detectors discussed in this paper provides helpful results for identification of particular diseases such as tooth decay, dental injury, cavities, abscess. And, some edge detector gives improvement in visibilities of objects (teeth) in the dental images. Dental X-rays are not sufficient for interpretation of diseases, area of missing teeth, caries-affected areas, root length for root canal treatment, and so on. Then, edge detection of dental X-ray images takes place to overcome the disadvantages of dental X-ray images because by the detection of edges in the image easy differentiation of objects takes place, which is very helpful for doctors in better interpretation of diseases for further diagnosis. As edge detection in dental images is very helpful, it should be more accurate so that the actual diseases are interpreted by doctor for diagnosis.

Table 2 Results of various edge detection techniques

References	Type of input image	Edge-detected image outcomes
[3]	Bitewing X-ray	Identification of tooth decay, supporting bones, and dental injury
[4]	Periapical X-ray	Improvement in visibility of individual tooth
[5]	Panoramic X-ray	Identification of cavities
[6]	Bitewing X-ray	Area of missing teeth
[8]	3-D teeth image	Measurement of tooth length for proper alignment
[9]	CBCT segment	3D reconstruction of teeth
[10]	Bitewing X-ray	Improvement in identification of tooth decay, supporting bones, and dental injury
[12]	Periapical X-ray	Diagnosis of abscess
[13]	Periapical X-ray	Improvement in endodontic working length measurement
[14]	Periapical X-ray	Caries-affected areas
[16]	CBCT segment	3D reconstruction of oval cavity
[18]	Panoramic X-ray	Dental age assessment
[19]	Bitewing X-ray	Improvement in visibility of individual tooth

The results of reviewed techniques have also detected some false edges which may lead to improper diagnosis of diseases, and in that case, a precise edge detection technique is required. This precision may be achieved by the use of general type-2 fuzzy logic system (GT2FLS) [28, 29] for edge detection because this approach deals with very high level of uncertainty in the problems, by the use of footprint of uncertainty (FOU) in its membership function which allows an improved solution of real-world uncertainty.

Fig. 12 ISEF method



References

1. Solanki C, Godfrey WW (2016) Technique for edge detection based on interval type-2 fuzzy logic with sobel filtering. In: IEEE transactions. Doi 978-1-5090-1987-8
2. Kaushik A, Mathpal PC, Sharma V (2014) Edge detection and level set active contour model for the segmentation of cavity present in dental X-ray images. Int J Comput Appl 96(9):0975–8887
3. Senthilkumaran N (2012) Edge detection for dental X-ray image segmentation using neural network approach. Int J Comput Sci Appl (TIJCSA), 1(7)
4. Ansingkar NP, Dhopeshwarkar MG (2014) Study and analysis of edge detection techniques for segmentation using dental radiograph. Int J Eng Comput Sci 3(9)
5. Na'am J, Harlan J, Madenda S, Wibowo EP (2016) The algorithm of image edge detection on panoramic dental X-ray using multiple morphological gradient (mMG) method. Int J Adv Sci Eng Sci Technol 6
6. Croock MS, Khudhur SD, Taqi AK (2016) Edge detection and features extraction for dental X-ray. Eng Tech J 34 Part (A)(13)

7. Lin PL, Huang P−W, Cho YS, Kuo C−H (2013) An automatic and effective tooth isolation method for dental radiographs. Opto−Electron https://doi.org/10.2478/s11772-012-0051-9
8. Saoji SU, Jaini P (2014) Line analyzer techniques for teeth using edge-based method and gray-based method. In: International conference on communication systems and network technologies. https://doi.org/10.1109/csnt.2014.186
9. Senthilkumaran N (2012) Genetic algorithm approach to edge detection for dental X-ray image segmentation. Int J Adv Res Comput Sci Electron Eng (IJARCSEE) 1(7)
10. Gayathri V, Menon HP (2014) Challenges in edge extraction of dental X-ray images using image processing algorithms—a review in (IJCSIT). Int J Comput Sci Inf Technol 5(4)
11. Pavaloiu I-B, Goga N, Vasilateanu A, Marin I, Ungar A, Patrascu I, Ilie C (2015) Neural network based edge detection for CBCT segmentation. IEEE 978-1-4673-7545-0
12. Mahant PM, Desai NP, Jain KR, Mahan MG (2015) Optimal edge detection method for diagnosis of abscess in dental radiograph. IJRSI II(II)
13. Padma Vasavi K, Udaya Kumar N, Madhavi Latha M, Krihna Rao EV An edge detection scheme for endodontic working length measurement in root canal treatment for succedaneous teeth in latest trends. Circ Syst Sig Process Autom Control. ISBN: 978-960-474-374-2
14. Solanki AJ (2016) Threshold selection in ISEF based identification of dental caries in decayed tooth. Int J Electron Electr Comput Syst (IJEECS) 5(5). ISSN 2348-117X
15. Trivedi DN, Shah N, Kothari AM (2016) Dental contour extraction & matching with label contouring using ISEF algorithm on DICOM images for human identification. Int J Latest Trends Eng Technol (IJLTET) 7(2)
16. Pavaloiu I-B, Goga N, Marin I, Vasilateanu A (2015) Automatic segmentation for 3D denta reconstruction. In: ICCCNT
17. Kamencay P, Zachariasova M, Hudec R, Benco M, Radil R (2014) 3D image reconstruction from 2D CT slices 3DTV-conference: the true vision—capture, transmission and display of 3D video (3DTVCON)
18. Razali MRM, Ahmad NS, Hassan R, Zaki ZM, Ismail W (2015) Sobel and Canny edges segmentations for the dental age assessment. IEEE. DOI 10.1109
19. Bhargavi K, Jyoth S (2016) An efficient fuzzy logic based edge detection algorithm. Int J Tech Res Appl 4(3)
20. Aborisade DO (2010) Fuzzy logic based digital image edge detection global. J Comput Sci Technol 10(14) (Ver. 1.0)
21. Senthilkumaran N (2012) Fuzzy logic approach to edge detection for dental X-ray image segmentation. (IJCSIT) Int J Comput Sci Inf Technol 3(5)
22. Tangel ML, Fatichah C, Yan F, Betancourt JP, Widyanto RM, Dong F, Hirota K (2013) Dental classification for periapical radiograph based on multiple fuzzy attribute. IEEE 978-1-4799-0348-1
23. Lai YH, Lin PL (2008) Effective segmentation for dental X-ray images using texture- based fuzzy inference system. LNCS 5259:936–947
24. Moynihan P, Petersen PE (2004) Diet, nutrition and the prevention of dental diseases. Public Health Nutr. https://doi.org/10.1079/phn2003589
25. Goryawala SN, Chavda P, Udhani S, Shukla D, Pathak S, Ojha R (2015) A survey on incidence of common dental problems among patients attending dentistry OPD at a tertiary care hospital from central Gujarat. Int J Res Med
26. Harris M, Eaton KA (2011) Discussion paper, dental hyginest and dental research: a developing scene OHDM 10(4)
27. Shivpuje BV, Sable GS (2016) A review on digital dental radiographic images for disease identification and classification. Int J Eng Res Appl 6(7) (Part -5):38–42. ISSN 2248-9622
28. Melin P, Gonzalez CI, Castro JR, Mendoza O, Castillo O (2013) Edge detection method for image processing based on generalized type-2. Fuzzy Logic IEEE. https://doi.org/10.1109/tfuzz.2013.2297159
29. Melin P, Gonzalez CI, Castro JR, Mendoza O, Castillo O (2016) General type-2 Fuzzy edge detector applied on face recognition system using neural networks. IEEE 978-1-5090-0626-7

Detection of Exudates and Removal of Optic Disk in Fundus Images Using Genetic Algorithm

K. Gayathri Devi, M. Dhivya and S. Preethi

Abstract Diabetic retinopathy is one of the serious and sight-threatening complications of diabetics. The main symptom of diabetic retinopathy is the presence of exudates that results in yellow flecks due to the fluid that has seeped out of damaged capillaries. This causes the tissue in the retina to distend, resulting in hazy or unclear vision. If they are left untreated, diabetic retinopathy can cause blindness. Hence, segmentation of exudates is vital process in retinal pathologies. The proposed work involves accurate segmentation of exudates from the retinal fundus images. Initially, K-means clustering is applied on the retinal images to separate the exudates and optic disk. Genetic algorithm is used for the accurate segmentation of the exudates in which the fitness function is calculated to perform crossover between the segmented images obtained from the K-means clustering segmentation. Before performing the mutation process, the grayscale image is converted into the RGB channels. These three-segmented channels are further combined by the mutation process to obtain the genetic algorithm output. High-intensity region is determined to be the exudates and the low intensity is said to be the optic disk. The elimination of the optic disk which has the same intensity as that of the exudates is performed using watershed segmentation. Finally, the parameter validation is done after the morphological operations. This method was implemented in 10 images downloaded from CHASE and STARE database and the accuracy has been improved to 94% compared with the existing approaches.

K. Gayathri Devi (✉) · M. Dhivya · S. Preethi
Electronics and Communication Engineering, Dr. N G P Institute of Technology,
Coimbatore, India
e-mail: gayathridevik@yahoo.com

M. Dhivya
e-mail: dhivya@drngpit.ac.in

S. Preethi
e-mail: preethi.s@drngpit.ac.in

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_129

1 Introduction

Diabetic retinopathy is an eye disease that causes abnormalities in the retina due to the complication of diabetes. The early identification of diabetic retinopathy is very much needed to save the vision and to provide treatment. Exudates are one of the symptoms of diabetic retinopathy. Diabetic retinopathy can be categorized into two stages such as non-proliferative stage and the proliferative stage [1]. Exudates identified in the non-proliferative stage are categorized as soft exudates and in the proliferative stage as hard exudates [2]. These exudates cause harm to the blood vessels in retina, and hence the capillaries in the vessels will leak some fluid and these will appear as shiny yellow–white dots with sharp borders called as exudates. Exudates can be identified by implementing image segmenting methods such as watershed algorithm, region-based segmentation methods, thresholding method, representation techniques, and machine learning methods [3, 4].

2 Existing Methods

Diabetic retinopathy is one of the major causes of blindness that can cause damage to blood vessels in the eye which will subsequently progress to the formation of lesion in the retina [1, 5]. Accurate exudates segmentation plays a vital role in diagnosis of diabetic retinopathy. Aqeel et al. [6] proposed an automated algorithm for the segmentation of exudates. For optimal results, texture and adaptive threshold algorithm have been incorporated for the accurate segmentation of exudates.

Pereira et al. [7] implemented a technique to extract the blood vessels in retinal fundus image for the automatic detection of diabetic retinopathy. The steps involved in this paper are preprocessing, feature extraction, ant colony optimization algorithm, and the post-processing. Two features are extracted from images and combined, in order to develop ant movement heuristics. The ant colony optimization is quite slow when compared to genetic algorithm and is efficient only in extraction of blood vessel. This proposed work does not consider the segmentation of exudates that is needed to diagnose diabetic retinopathy.

Soares et al. [8] suggested a technique to localize the optic disk (OD) and the importance of diagnosing diabetic retinopathy. The localization of the OD was done by cumulative sum field technique. The algorithm reveals to be reliable and efficient. Yu et al. [9] enhanced the detection of optic disk using directional filtering technique and level set method which requires the placement of seed point to localize the optic disk.

Radha et al. [10] developed a process to detect the exudates and to classify the retina to be normal or abnormal. The combination of multi-structure morphological process and *K*-means clustering technique is used effectively for retinal vessel and exudates detection. Exact detection of the condition of a retina whether it is normal or abnormal was determined successfully. For the abnormal retina, the exudates were detected by extracting the blood vessels using plane separation method. The features were extracted by applying discrete wavelet transform (DWT) and energy feature coefficients were trained with the probabilistic neural network for classifying the exudates. Morphological operations are applied on segmented image for smoothening the exudates part.

The supervised learning approach based on the multilayer perceptron (MLP) was used by Kusakunniran et al. [11] to determine the elementary seeds for the segmentation of hard exudates. In the preprocessing step, color transfer has been applied to normalize the color information to reduce the variation of color that will exist among hard exudates in different retinal images. Then, the supervised and unsupervised techniques are applied in combination to segment the exudates.

Esther et al. [12] emphasized on the segmentation of optic disk using watershed segmentation. The retinal fundus image is used by the ophthalmologists for analyzing the important features of the eye like the geometrical features of optic disk, the anterior segment of the retina, and any abnormal growth of any region in the eye. Finding the position of the optical disk in the retina plays a vital role in the diagnosis and differentiation of the exudates from the optical disk. Then, the exact contours are found using watershed transformation. The methodology for the extraction of the OD contour is mainly based on mathematical morphology along with thresholding and watershed transformation.

Vimala et al. [13] proposed an automatic and efficient technique for the identification of OD and exudates. The preprocessing of retinal images was implemented to transform the RGB image into L*A*B color space. The segmentation of OD in the preprocessed LAB image was performed using line operator and fuzzy C means clustering technique. The extraction of exudates was done by applying *K*-means clustering with a predefined threshold and number of clusters, and finally the classification was done using support vector machine.

Hole et al. [14] enhanced the detection of exudates by applying genetic algorithm to improve the accuracy, reduce the computational time, and for its easier diagnosis. These various approaches based on genetic algorithm were suggested to produce an image with good and natural contrast as the distribution of intensity level of the pixels in a fundus image is within a small range. Jelinek et al. [15] proposed a technique in which fusion technique was applied to obtain accurate segmentation.

Thus, based on the above literature review, the techniques that were implemented to improve the accuracy in segmentation exudates from the retinal fundus images are

- Initially, *K*-means clustering is applied on the retinal images to separate the exudates and optic disk.
- Genetic algorithm (GA) is applied to the clustered output in which the fitness function is calculated and then ranked to obtain the best case, average case, and worst case to improve the accuracy of segmentation.
- The elimination of the optic disk which has the same intensity as that of the exudates is performed using watershed segmentation.

3 Proposed Methods

The flow diagram of the proposed method is shown in Fig. 1, the input image was taken from the CHASE and STARE database for the segmentation of exudates. The intensity value of the pixel in the retinal images will range from 0 to 255 (for each R, G and B planes). The preprocessing was performed by converting RGB channel into the L*A*B channels and later it is converted to gray scale.

Segmentation is done using K-means clustering algorithm in which the centroids are calculated and the mean values are initialized. Genetic algorithm is used for the accurate segmentation of the exudates in which the fitness function is calculated to perform crossover and mutation in which the best case, average case, and worst case are determined. Crossover is performed between the segmented images obtained from the K-means clustering segmentation. Before performing the mutation process, the grayscale image is converted into the RGB channels. These three-segmented channels are further combined by the mutation process to obtain the genetic algorithm output.

The optic disk is eliminated using watershed segmentation which is performed only for the grayscale images. In watershed segmentation, gradient magnitude is used as a segmentation function. Background markers are used for eliminating the optic disk.

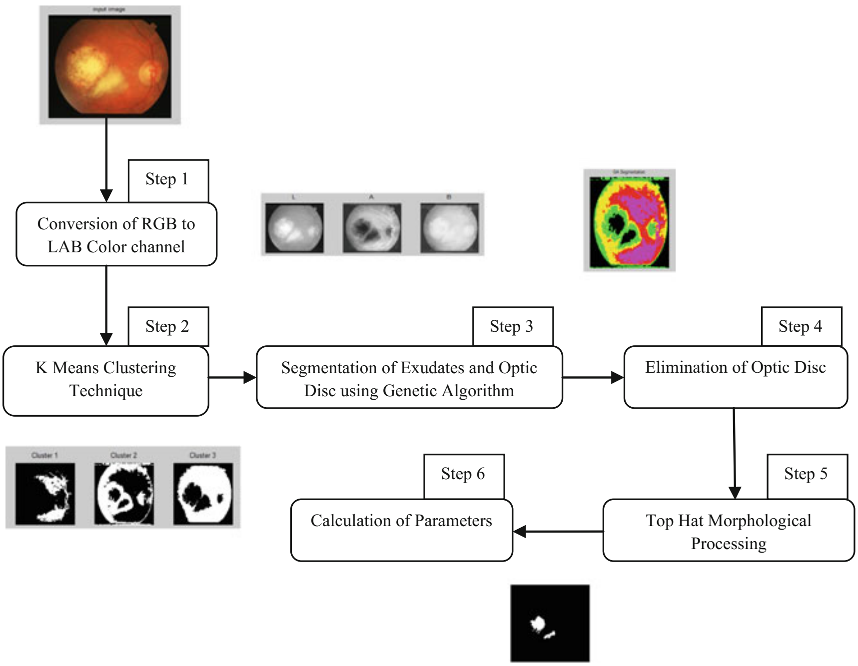


Fig. 1 Flow diagram of the proposed method

3.1 Preprocessing

The input image consisting of 700 × 605 pixels stored in TIFF format are converted into L*a*b channels using color space conversion and the preprocessed image is considered for the segmentation process. The input images are resized to focus on the region of interest and to reduce the execution time. The input image is shown in Fig. 2a and the preprocessed image, that is, resized image for the consequent process is shown in Fig. 2b.

The RGB image is converted into L*a*b* color space as it is the most exact technique for the representation of color information and can be applied to any portable device. LAB color space consists of a luminosity layer L* and two chromaticity layers a* and b* which provides information of the axis where the red–green or blue–yellow exist as the predominant color in the fundus images are red and yellow colors. The major color information of the image lies in the a* and b* layer. The flow diagram shown in Fig. 3 explains the conversion of input fundus images into L*a*b channels.

The input fundus image in RGB format is converted into grayscale format for the application of watershed segmentation for the elimination of optic disk which is shown in Fig. 4a. The LAB channel images converted into grayscale format are shown in Fig. 4b. The preprocessed image is subjected next for segmentation process, a* channel has been considered as the localized information of red which is important for further identification of optic disk and extraction of exudates.

The extracted A channel image components are grouped into a set of three clusters using K-means clustering algorithm. The main aim of using K-means clustering is to partition the image component into K clusters so that each observation will belong to one cluster to the nearest centroid that has been initialized. K-means clustering algorithm makes sure that the clusters are constructed based on comparable spatial extent and the expectation-maximization mechanism allows clusters to have different shapes.

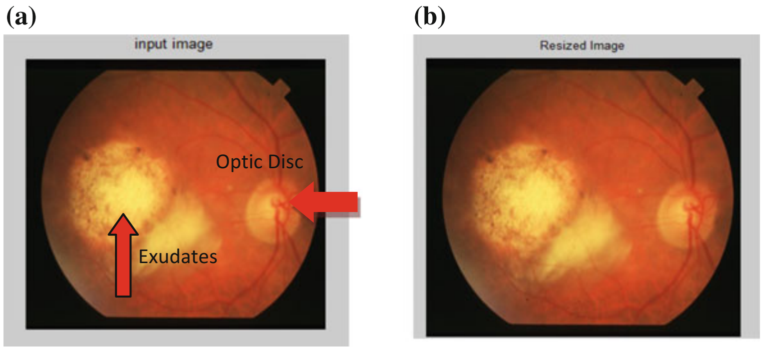


Fig. 2 **a** Input retinal image with exudates. **b** Resized image

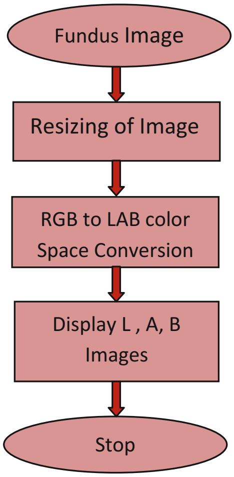


Fig. 3 Flow diagram of plane separation

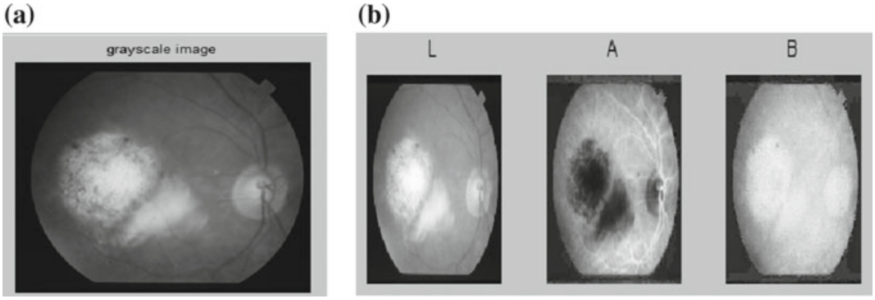


Fig. 4 **a** Grayscale image. **b** L*a*b Channels of input image

By continuous iteration of K-means algorithm with a range of K values, the cluster data sets are grouped and the results are compared. The cluster centroid and the mean distance between the data points are the two main metrics used for comparing the results across different K values. Then, the algorithm finds the relevant cluster and data set are labelled for particular pre-chosen cluster $K = 3$. The algorithm to implement the K-means clustering is composed of the following steps:

(i) Initialization process:
Initialization of K points for representing the objects to be clustered.

(ii) Grouping into clusters:
The pixels that are nearby to the initialized centroid are identified by Euclidean distance measure parameter and the grouping of pixels in the image is performed.

(iii) Optimization of centroid of clusters:

Fig. 5 Segmented output

When all pixels have been assigned to any one cluster, the positions of the *K* centroids are recalculated again.

(iv) Reiteration process:
Steps (ii) and (iii) are repeated until the centroids can no longer move. This produces the separation of the objects into clusters groups.

The entire process can be summarized in following steps:

1. The input RGB retinal fundus image is converted into L*A*B channels and A* channel output is considered for further subsequent steps.
2. Apply *K*-means clustering algorithm based on Euclidean distance metric to group the pixels in the image into three clusters.
3. Every pixel in the image is labeled and it returns an index corresponding to a cluster. Every pixel in the image is labeled with its cluster index.
4. The output of *K*-means clustering for the three clusters segmented is shown in Fig. 5. Thus, the cluster 3 output shown in Fig. 5c containing optic disk and exudates are further considered for the segmentation of exudates.

3.2 Segmentation of Exudates and Optic Disk Using Genetic Algorithm

The clustered outputs are processed with genetic algorithm for optimization and to improve the accuracy of the segmented output. Genetic algorithm (GA) is used to obtain the accurate optimization output for the segmentation of exudates. In addition to that, this method is suitable for all retinal images with both exudates and the optic disk which is used for the diagnosis of diabetic retinopathy. Figure 6 highlights the steps that are performed in genetic algorithm. A typical genetic algorithm requires the following:

1. Representation of the clustered output solution to process in GA domain.
2. Fitness function to evaluate the solution domain.

Each candidate solution is represented as an array of bits. Arrays of other types and structures are used in essentially the same way. To facilitate the simple crossover

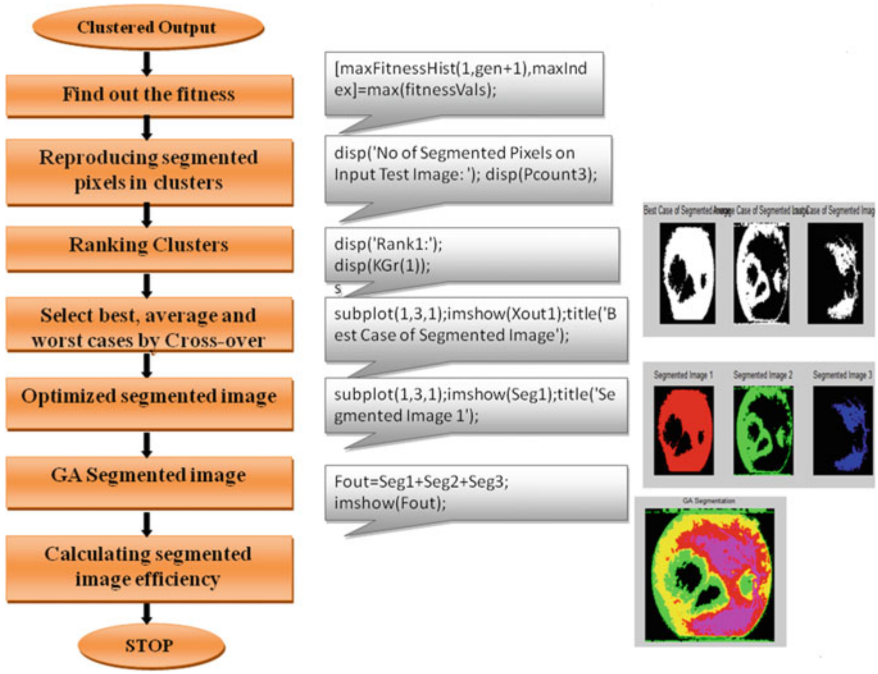


Fig. 6 Flowchart to perform genetic algorithm

operations, the parts in the images are aligned according to their fixed sizes and this main property makes the genetic representations more convenient.

The important steps that are to be performed are as follows:

Step 1 Encoding the problem in a binary string.
Step 2 Random population generation.
Step 3 Fitness calculation of each solution.
Step 4 Pairs of parent strings are selected based on fitness.
Step 5 New strings are generated with crossover and mutation process until a new population has been produced.
Step 6 Repeat steps 2–5 until satisfying solution is obtained.

Initialization Process

The genetic algorithm requires individual which should be represented by strings, here the cluster 3 output is encoded into integer strings. The region containing exudates (indicated by red arrow) and optic disk (indicated by blue arrow) are only black in color and other regions are white as shown in Fig. 5c. The population size depends on the number of pixels in exudates and optic disk region. Here, generic representation is "seeded" in these areas of interest where optimal solutions are likely to be found. Here, in our example image there are three regions ($r1$, $r2$, $r3$), two exudates and one optic disk region. A number of chromosome $\alpha = \alpha 1, \alpha 2, \alpha 3$ are

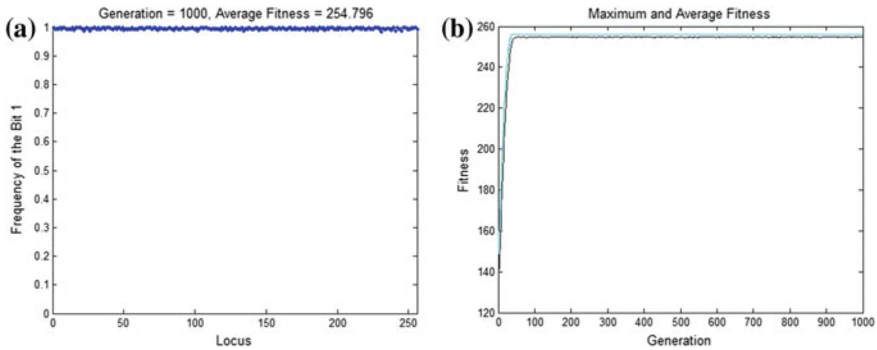


Fig. 7 **a** Fitness generation calculation. **b** Graph between generation and fitness

encoded as integer strings. The number of chromosomes will be equal to the number of segmented region of interest.

Selection Process

During each successive generation, a portion of the existing population should be selected in order to proceed further to breed a new generation. The fitness function is based on the region of selection in the cluster 3. The properties that are considered for the fitness function are size of the region and intensity of pixels which decide the similarity and the variation between regions. The fitness function can be calculated as the inverse of product of the variance and the size of each region. The best individuals are selected based on the above fitness function.

Roulette wheel selection operators are used for the selection in which the individuals with the best fitness function occupies a larger portion in the wheel and vice versa. Individual solutions are selected through a *fitness-based* process (Fig. 7a), where fitter solutions (as measured by a fitness function) are typically more likely to be selected. Graph between generation and fitness id obtained are shown in Fig. 7b.

Genetic Operators

The genetic operators like crossover and mutation are used to produce a second generation population of solutions.

After the selection of the chromosome, two points crossover operator can be applied to produce two offsprings. The length where the crossover should occur depends on the variation of intensity of pixels in the portioned region. This new solution created, typically shares many of the characteristics of its parents. Consider the representation of two parents which is shown in Fig. 8. The genetic representation of two exudates region as parent 1 and parent 2 is divided into 10 cells where each cell represents a gene. The crossover may occur between R3 and R5 of parent 1 with R3 and R5 of parent 2 and R8 and R9 of parent 1 with R8 and R9 of parent 2.

For each new child, new parents are selected and the process continues until a new population of solutions of appropriate size is generated. Although reproduction

Parent 1

R1	R2	R3	R4	R5	R6	R7	R8	R9	R10

Parent 2

R1	R2	R3	R4	R5	R6	R7	R8	R9	R10

Fig. 8 Selected parents for crossover operation

Fig. 9 Segmented images

methods that are based on the use of two parents are more "biology inspired", some research suggests that more than two "parents" generate higher quality chromosomes.

Thus, the next generation population of chromosomes that is different from the initial generation is developed. This will increase the average fitness for the population, since only the best organisms from the first generation are selected for breeding, along with a small proportion of less fit solutions. These less fit solutions ensure genetic diversity within the genetic pool of the parents, and therefore ensure the genetic diversity of the subsequent generation of children.

Termination of GA

This generational process will be repeated until a termination condition has been reached. There are some conditions that are to be considered. Common terminating conditions are as follows:

- A solution should be found that satisfies the minimum criteria.
- Fixed number of generations has been reached.
- Allocated budget (computation time/money) is reached.
- The highest ranking solution's fitness is reaching or has reached a plateau such that successive iterations no longer produce better results.
- Manual inspection.
- Combinations of the above.

Best case, worst case, and average cases are obtained and shown in Fig. 9. These segmented outputs will be considered for the termination process to find the final GA segmentation output which is shown in Fig. 10.

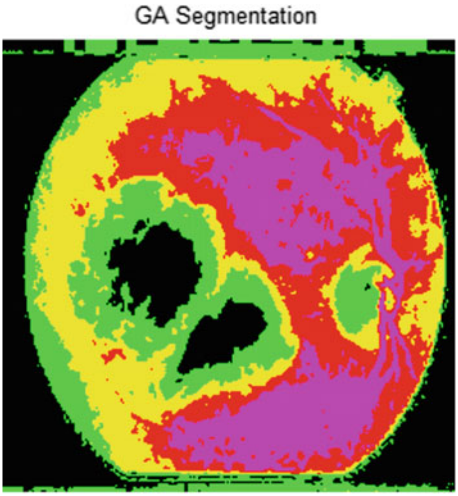


Fig. 10 GA segmentation output

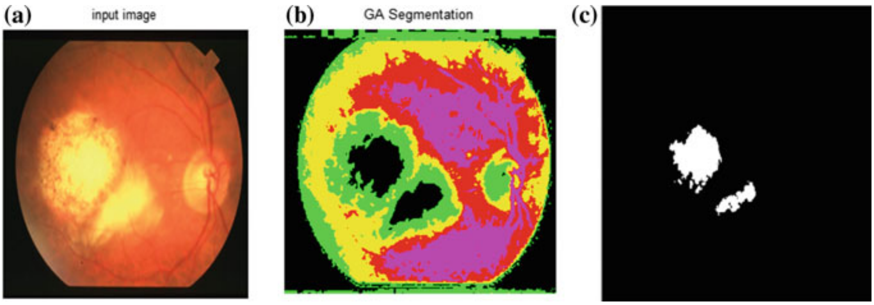


Fig. 11 **a** Input image. **b** GA segmented output containing exudates and OD. **c** Elimination of OD

3.3 Elimination of Optic Disk

Eigenvector-based OD detection is implemented to calculate the covariance matrix of the database images and the disk region is constructed from the eigenvectors. The region with minimum distance is grouped as optical disk. This work is used to distinguish the exudates and the optic disk in an image. This approach is computationally inexpensive and highly robust. Morphological operators are determined for the OD detection. Dilation, top hat transformation operators are used for the elimination of OD. The time taken by this algorithm is also very low. Watershed segmentation technique is used for average filtering and contrast enhancement as preprocessing steps. This is further processed to obtain the modified gradient vector flow algorithm for OD segmentation. Figure 11a shows the input image and 11c reveals the elimination of OD and presence of exudates.

4 Results and Discussion

The classification (or prediction) data is divided into two different classes, positives (P) and negatives (N). This classification produces four types of outcomes—true positive, true negative, false positive, and false negative.

- True positive (TP): Exudates correctly identified as exudates.
- False positive (FP): Non-exudates pixels incorrectly identified as exudates.
- True negative (TN): Non-exudates pixels correctly identified as non-exudates pixels itself.
- False negative (FN): Exudates incorrectly identified as non-exudates.

Accuracy

Accuracy is defined and calculated using Eq. 1 as the ratio of sum of TP and TN (correct identification of exudates) divided by the total number of the images.

$$\mathrm{ACC} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FN} + \mathrm{FP}} \tag{1}$$

Sensitivity

Sensitivity (SN) will provide us with the details of identification of exudates region. Sensitivity is calculated using Eq. 2 which is defined as the ratio of correctly predicted exudates (TP) to the total number of TP and FN.

$$\mathrm{SN} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{2}$$

Specificity

Specificity is also called the true negative rate (TNR) and is calculated as given in Eq. 3. Specificity gives us the correct identification of non-exudates region (TNR, expressed as a percentage).

$$\mathrm{SP} = \frac{\mathrm{TN}}{\mathrm{TN} + \mathrm{FP}} \tag{3}$$

The proposed method is compared with existing method in terms of the algorithm used and three parameters, namely, accuracy, sensitivity, and specificity and has also been utilized to evaluate the improvement in the performance of the proposed method as shown in Table 1 and the outputs for different test images are shown in Fig. 12. The proposed technique shows an improvement in all the parameters.

Table 1 Comparison of existing and proposed method parameters

Parameter	Existing technique [16]	Proposed technique
Algorithm used	Adaptive threshold algorithm (%)	Genetic algorithm (%)
Accuracy	93.69	94
Sensitivity	90.42	93
Specificity	94.60	95

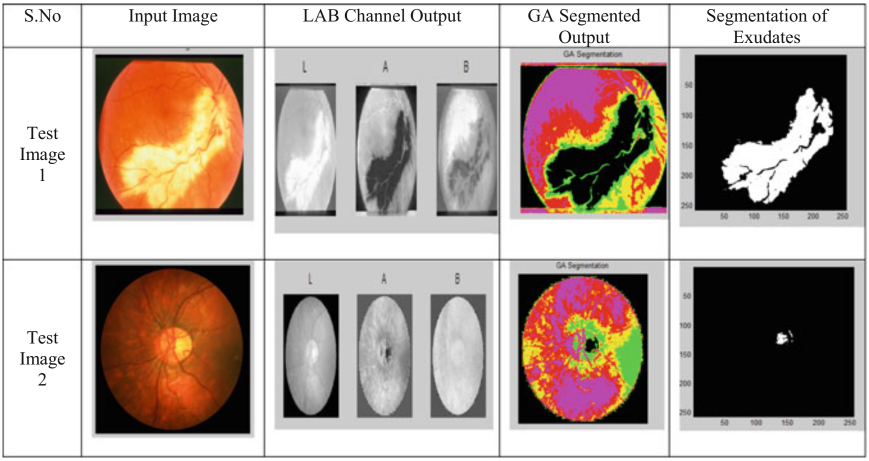


Fig. 12 Segmentation of exudates for the different test images

5 Conclusion

Diabetic retinopathy is a disease caused to a diabetic patient, it is progressive, and hence this requires continuous monitoring of the abnormalities in the eye which may lead to loss of vision. A novel method that aims to improve the accuracy in the segmentation of exudates to help the ophthalmologists for the identification and the computer-aided diagnosis of diabetic retinopathy has been proposed and analyzed in this work.

References

1. Antal B, Hajdu A (2012) An ensemble-based system for microaneurysm detection and diabetic retinopathy grading. IEEE Trans Biomed Eng 59(6):1720–1726
2. Deshmukh AV, Patil TG, Patankar SS, Kulkarni JV (2015) Features based classification of hard exudates in retinal images. In: 2015 international conference on advances in computing, communications and informatics (ICACCI), pp 1652–1655

3. Pires R, Avila S, Jelinek HF, Wainer J, Valle E, Rocha A (2014) Automatic diabetic retinopathy detection using bossa nova representation. In: 2014 36th annual international conference of the IEEE engineering in medicine and biology society (EMBC), pp 146–149
4. Roychowdhury S, Koozekanani DD, Parhi KK (2014) Dream: diabetic retinopathy analysis using machine learning. IEEE J Biomed Health Inform 18(5):1717–1728
5. Esmaeili M, Rabbani H, Dehnavi AM, Dehghani A (2012) Automatic detection of exudates and optic disk in retinal images using curvelet transform. IET Image Proc 6(7):1005–1013
6. Aqeel, AF, Ganesan S (2014) Automated algorithm for retinal image exudates and Drusens detection, segmentation, and measurement. In: 2014 IEEE international conference on electro/information technology (EIT), pp 206–215
7. Pereira C, Gonçalves L, Ferreira M (2015) Exudate segmentation in fundus images using ant colony optimization approach. Inf Sci 296:14–24
8. Soares I, Castelo-Branco M, Pinheiro AM (2016) Optic disc localization in retinal images based on cumulative sum fields. IEEE J Biomed Health Inform 20(2):574–585
9. Yu H, Barriga ES, Agurto C, Echegaray S, Pattichis MS, Bauman W, Soliz P (2012) Fast localization and segmentation of optic disk in retinal images using directional matched filtering and level sets. IEEE Trans Inf Technol Biomed 16(4):644–657
10. Radha R, Lakshman B (2013) Retinal image analysis using morphological process and clustering technique. Sig Image Process 4(6):55
11. Kusakunniran W, Wu Q, Ritthipravat P, Zhang J (2018) Hard exudates segmentation based on learned initial seeds and iterative graph cut. Comput Methods Programs Biomed 158:173–183
12. Esther JJJ, Sophia SG (2014) Detecting optic disc in digital fundus images using stochastic watershed transformation. IJREAT Int J Res Eng Adv Technol 2(1)
13. Vimala GAG, Mohideen SK (2013) Automatic detection of optic disk and exudate from retinal images using clustering algorithm. In: 2013 7th international conference on intelligent systems and control (ISCO), pp 280–284
14. Hole KR, Gulhane VS, Shellokar ND (2013) Application of genetic algorithm for image enhancement and segmentation. Int J Adv Res Comput Eng Technol (IJARCET) 2(4):1342
15. Jelinek HF, Pires R, Padilha R, Goldenstein S, Wainer J, Bossomaier T, Rocha A (2012) Data fusion for multi-lesion diabetic retinopathy detection. In: 2012 25th international symposium on computer-based medical systems (CBMS), pp 1–4
16. Wisaeng K, Hiransakolwong N, Pothiruk E (2015) Automatic detection of exudates in retinal images based on threshold moving average models. Biophysics 60(2):288–297

Analysis of Feature Ranking Methods on X-Ray Images

H. Roopa and T. Asha

Abstract Mycobacterium causes an infectious disease called tuberculosis which can be diagnosed by its various symptoms like fever, cough, etc. Tuberculosis can also be analyzed by understanding the chest X-ray of the patient which is revealed by an expert physician. The chest X-ray image contains texture and shape-based features which are extracted from X-ray image using image processing concepts. This paper presents implementation of various feature weighting methods on the extracted features of X-ray images. These feature weighting methods are analyzed using linear regression model and Linear Discriminant Analysis (LDA) model. The performance of various feature weighting methods is compared and found that the accuracy of weights by PCA using linear regression model is 98.75% which is better than other methods.

1 Introduction

Mycobacterium causes infectious disease tuberculosis (TB) [1] which usually affects lungs. When an infected person coughs, TB is transmitted from one person to another through air. TB affects lungs but can also appear in other parts of body like spine, bones, brain, and kidney. Chest X-ray findings, fever, loss of weight, and coughs are some symptoms of TB. To raise awareness about TB, world TB day is recognized on March 24 to support prevention of TB. The level of TB in all country is monitored by World Health Organization (WHO) [2]. Every year more than 9 million people develop TB and 1.5 million die from the disease.

Features represent information in an image. To understand the information present in an image, it must be analyzed and measured in various angles to get relevant information in a particular domain like medical image, satellite images, etc., so,

H. Roopa (✉) · T. Asha
Department of C. S. E, Bangalore Institute of Technology, K. R. Road, Bangalore, India
e-mail: roopatejas@gmail.com

T. Asha
e-mail: asha.masthi@gmail.com

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_130

feature extraction plays an important role in analyzing medical image like TB image. Feature ranking is one of the feature selection methods where the features are ranked using certain defined measures.

Linear Regression (LR) model is a method to find a relationship between one dependent variable and series of changing independent variables by the straight line that represents the linear equation of the observed data. Linear Discriminant Analysis (LDA) separates two or more classes to find the best linear combination of features.

The aim of this paper is to analyze extracted features of chest X-ray images. Then, these features are ranked using different feature weighting methods like Relief, Principle Component Analysis (PCA), Support Vector Machine (SVM), etc. and examined using linear regression model and LDA model for classifying TB disease.

The paper is organized as follows: a brief overview of related work is given in Sects. 2 and 3 describes method for extracting and ranking features of X-ray image, and analyze these by applying Linear Regression (LR) model and LDA model. Section 4 gives details of experimental illustration and results of various ranking methods are explained in Sect. 5. Finally, Sect. 6 concludes the paper.

2 Related Work

For medical image analysis, Perner et al. [3] developed data mining and image processing tool using image mining concepts. Descriptions of list of attributes as given by experts are stored in a database and then a classification technique decision tree induction tree is applied to this to extract expert knowledge. This tool was used for various applications like breast MRI data, etc. Asha et al. [4] used Association Rule Mining (ARM) techniques on TB data sets to improve TB disease prediction. The symptoms of TB were considered and many descriptive rules were written and these were combined with an association classification technique used for predicting TB.

Zou et al. [5] proposed a Max-Relevance-Max-Distance (MRMD) feature extraction method, which provides stabilized feature selection method using new measure index for ranking the features that improve the efficiency of classification. Bravi et al. [6] used concave approximation of zero-norm function for ranking the features and evaluated it using SVMs with Gaussian kernel. Chung et al. [7] used feature selection and Taguchi genetic algorithm together on DNA microarray data and then the performance was evaluated using KNN with Leave-One-Out Cross-Validation (LOOCV).

Razmjoo et al. [8] developed an efficient way to rank features in an incremental manner which can be used in classification methods. Yihui et al. [9] extracted wavelet features from microarray data, then SVM was used for classification. Zyout et al. [10] extracted textual pattern from mammogram images and Particle Swarm Optimization

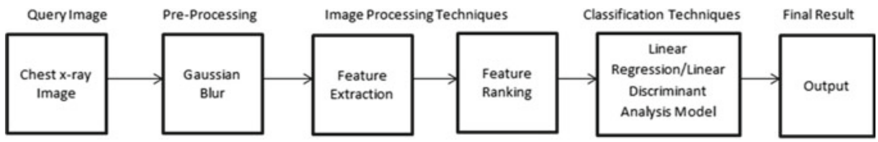


Fig. 1 Feature extraction and feature ranking of X-ray image

(PSO) was applied to select the most discriminative features, then SVM was used for classification. City block distance measure was used by Roopa et al. [11] for segmenting chest X-ray image which helps in diagnosis of TB.

3 Proposed Methodology

Features which are relevant and informative with respect to TB disease should be considered. The X-ray image must first be preprocessed and then important features are extracted from the affected region of X-ray using image processing methods. These features are ranked using various methods and evaluated using LR and LDA models. The feature extraction and ranking of an X-ray image is shown in Fig. 1.

The steps involved in extracting and ranking features from X-ray image are

1. X-ray image is taken as an input.
2. Preprocessing:
   Noise and redundant data in the X-ray image are removed using Gaussian blur filter method.
3. Feature extraction:

Geometric features and texture-based features [12] can be used to measure the characteristics of TB. Both the shape descriptors and texture descriptors were extracted from X-ray image. Extracting as many features from the region of interest of TB is one of the concerns in this work which is done by applying roipoly () function using MATLAB software.

Shape-based features were used to measure the region of interest in a TB image. The statistics like area, perimeter, coordinates of region centroid, major axis length, minor axis length, eccentricity, orientation, etc. are some of the characteristics of shape-based feature extraction which is obtained by analyzing external boundary of the X-ray image. These features are obtained using function called as regionprops ().

Texture descriptors proposed by Haralick [13] defines 14 statistics that can be calculated from the co-occurrence matrix of the image. Texture feature extraction refers to surface characteristics and appearance of an object in an image. Entropy can be found using entropy (), graycrops () function can be used to extract homogeneity, contrast, energy, correlation, numel (UL) where UL represents uniformity, mean, standard deviation can be calculated using SD = sqrt(VR), where VR represents the variance and skewness using function skewness(). All these feature values were extracted from chest X-ray images.

4. Feature ranking:
   i. Weight by PCA:
   The attribute weight of TB data is obtained using component created by PCA. The attribute weight reflects the importance of attributes with respect to class attribute.
   ii. Weight by chi squared:
   For each attribute of TB data, the value of chi-squared statistic with respect to class attribute is calculated to obtain the weight of the attribute.
   iii. Weight by correlation:
   The weight of feature of TB data is calculated by computing value of correlation for each feature of TB data with respect to class attribute. The attribute weight is obtained by absolute value of correlation.
   iv. Weight by information gain:
   The attribute weight of TB data is obtained by calculating each attribute information gain with respect to class attribute.
   v. Weight by relief:
   The features of TB data values are distinguished between instances of same and different classes that are near each other to obtain weights of features.
   vi. Weight by SVM:
   The attribute weight of TB data is obtained by coefficients of hyperplane which is calculated by SVM with respect to class attribute.

5. Classification model:
   Linear regression model is applied on ranked features. Here, features are selected using M5 prime method by eliminating collinear features with a minimum tolerance of 0.05. Then, LDA model is applied to evaluate the ranked features and the results were compared with LR model.
6. Final result:
   Features obtained from X-ray images are classified as affected or normal image using LR and LDA model and the performance of feature weighting methods was evaluated using tenfold cross-validation technique.

4 Experimental Illustration

Consider TB-affected image, initially, the image is preprocessed to remove noise using Gaussian function. Then, texture- and shape-based features are extracted using MATLAB based on image processing concepts. The results of feature extraction method of TB X-ray image are illustrated in Figs. 2, 3, and 4, respectively.

The affected region of TB image is marked and then the marked region is masked to extract shape-based and texture-based feature represented in Figs. 3 and 4, respec-

Fig. 2 TB image

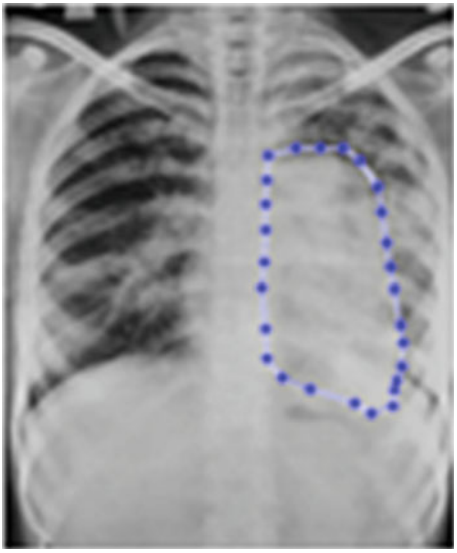


Fig. 3 Marked image

tively. Using the image processing methods, texture-based and shape-based feature values extracted from the masked image of TB are stored in a file. Table 1 gives information of extracted values of a single TB image.

For the implementation, we have considered 47 TB-affected images and 30 normal images. Features from X-ray images were extracted using MATLAB software and feature weighting and evaluation using classification model were performed using rapid miner software.

Fig. 4 Masked image

Table 1 Extracted feature values of a single TB image

Entropy	2.0954
Skewness	0.2724
SNR	6.1205
Homogeneity	0.9964
Contrast	0.16
Energy	0.4991
Correlation	0.9933
Total mean	110.4313
Variance	110.0112
Standard deviation	10.4886
Uniformity	111
Area	1744
Perimeter	172.625
Centroid	193.9060 218.2534
Major axis	65.7998
Minor axis	36.1286
Eccentricity	0.8358

All extracted feature values were used as an input for weighting the features using various feature ranking methods, and these features were evaluated using LR and LDA models separately to classify the X-ray image as affected or normal. The performance of the classification model was analyzed using tenfold cross-validation technique.

Table 2 Comparison results of feature ranking methods using LR and LDA models

Classification accuracy		
Feature ranking methods	Linear regression model	Linear discriminant analysis model
Weight by PCA	98.75	93.57
Weight by chi squared	96.07	90.8
Weight by correlation	96.07	90.8
Weight by information gain	96.07	85.8
Weight by relief	96.07	82.14
Weight by SVM	94.82	78.2

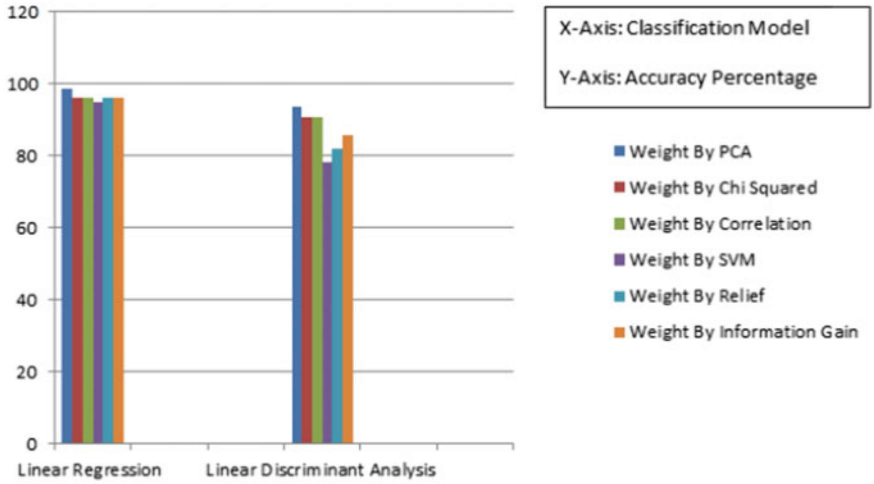


Fig. 5 Comparison of feature ranking methods using LR and LDA models

5 Results

Wrapper feature selection method is applied on extracted features of TB data. These features are ranked using various ranking methods and evaluated using LR and LDA models. The performances of these models are shown in Table 2 and Fig. 5, respectively.

The accuracy using weight by PCA of LR model is good when compared to LDA model, and also with respect to other feature weighting methods as shown in Table 2 and Fig. 5, respectively.

The implementation snapshots of feature weighting by PCA and SVM are given in Figs. 6, 7, 8, and 9, respectively.

```
PerformanceVector:
accuracy: 93.57% +/- 10.20% (mikro: 93.51%)
ConfusionMatrix:
True:   Affected        Normal
Affected:       47      4
Normal: 1       25
precision: 96.15% (positive class: Normal)
ConfusionMatrix:
True:   Affected        Normal
Affected:       47      4
Normal: 1       25
recall: 81.67% +/- 32.02% (mikro: 86.21%) (positive class: Normal)
ConfusionMatrix:
True:   Affected        Normal
Affected:       47      4
Normal: 1       25
AUC (optimistic): 1.000 +/- 0.000 (mikro: 1.000) (positive class: Normal)
AUC: 0.500 +/- 0.000 (mikro: 0.500) (positive class: Normal)
AUC (pessimistic): 0.000 +/- 0.000 (mikro: 0.000) (positive class: Normal)
```

Fig. 6 Performance of feature weighting by PCA using LDA model

```
PerformanceVector:
accuracy: 98.75% +/- 3.75% (mikro: 98.70%)
ConfusionMatrix:
True:   Affected        Normal
Affected:       47      0
Normal: 1       29
precision: 96.67% +/- 10.00% (mikro: 96.67%) (positive class: Normal)
ConfusionMatrix:
True:   Affected        Normal
Affected:       47      0
Normal: 1       29
recall: 100.00% +/- 0.00% (mikro: 100.00%) (positive class: Normal)
ConfusionMatrix:
True:   Affected        Normal
Affected:       47      0
Normal: 1       29
AUC (optimistic): 0.983 +/- 0.050 (mikro: 0.983) (positive class: Normal)
AUC: 0.983 +/- 0.050 (mikro: 0.983) (positive class: Normal)
AUC (pessimistic): 0.983 +/- 0.050 (mikro: 0.983) (positive class: Normal)
```

Fig. 7 Performance of feature weighting by PCA using LR model

PerformanceVector

```
PerformanceVector:
accuracy: 78.21% +/- 23.10% (mikro: 77.92%)
ConfusionMatrix:
True:   Affected        Normal
Affected:       41      10
Normal: 7       19
precision: 73.08% (positive class: Normal)
ConfusionMatrix:
True:   Affected        Normal
Affected:       41      10
Normal: 7       19
recall: 60.83% +/- 41.84% (mikro: 65.52%) (positive class: Normal)
ConfusionMatrix:
True:   Affected        Normal
Affected:       41      10
Normal: 7       19
AUC (optimistic): 1.000 +/- 0.000 (mikro: 1.000) (positive class: Normal)
AUC: 0.500 +/- 0.000 (mikro: 0.500) (positive class: Normal)
AUC (pessimistic): 0.000 +/- 0.000 (mikro: 0.000) (positive class: Normal)
```

Fig. 8 Performance of feature weighting by SVM using LDA model

6 Conclusion

The proposed work uses image processing concepts to extract relevant and important features from a chest X-ray image to diagnose whether a person is TB infected or not, by preprocessing data and feature extraction process. Then, wrapper feature selection methods are implemented on these extracted features and ranked using various methods. These features are classified using linear regression and LDA classification model to analyze the TB disease. The performance of feature weighting by PCA using LR model is examined and found that the accuracy is 98.75% which is better when compared to the accuracy of other feature weighting methods. Future work is to consider more images and carry out the implementation on this huge data set.

PerformanceVector

```
PerformanceVector:
accuracy: 94.82% +/- 6.36% (mikro: 94.81%)
ConfusionMatrix:
True:   Affected        Normal
Affected:       47      3
Normal: 1       26
precision: 96.30% (positive class: Normal)
ConfusionMatrix:
True:   Affected        Normal
Affected:       47      3
Normal: 1       26
recall: 84.17% +/- 30.38% (mikro: 89.66%) (positive class: Normal)
ConfusionMatrix:
True:   Affected        Normal
Affected:       47      3
Normal: 1       26
AUC (optimistic): 0.973 +/- 0.059 (mikro: 0.973) (positive class: Normal)
AUC: 0.973 +/- 0.059 (mikro: 0.973) (positive class: Normal)
AUC (pessimistic): 0.973 +/- 0.059 (mikro: 0.973) (positive class: Normal)
```

Fig. 9 Performance of feature weighting by SVM using LR model

References

1. Asha T, Murthy KNB, Natarajan S (2012) Data mining techniques in the diagnosis of tuberculosis. INTECH open access Publisher
2. Global Tuberculosis Report 2015-http://apps.who.int/iris/bitstream/10665/191102/1/9789241565059_eng.pdf
3. Perner P (2002) Image mining: issues, framework, a generic tool and its application to medical-image diagnosis. Eng Appl Artif Intell 15(2):205–216
4. Asha T, Natarajan S, Murthy KNB (2011) A study of associative classifiers with different rule evaluation measures for tuberculosis prediction. IJCA special issue on “Artificial intelligence techniques—novel approaches & practical applications” AIT
5. Zou Q, Zeng J, Cao L, Ji R (2016) A novel features ranking metric with application to scalable visual and bioinformatics data classification. Neurocomputing 173:346–54
6. Bravi L, Piccialli V, Sciandrone M (2017) An optimization-based method for feature ranking in nonlinear regression problems. IEEE Trans Neural Netw Learn Syst 28(4):1005–1010
7. Chuang L-Y et al (2011) A hybrid feature selection method for DNA microarray data. Comput Biol Med 41(4):228–237
8. Razmjoo A, Xanthopoulos P, Zheng QP (2017) Online feature importance ranking based on sensitivity analysis. Expert Syst Appl 85:397–406
9. Liu Y (2009) Wavelet feature extraction for high-dimensional microarray data. Neurocomputing 72(4):985–990
10. Zyout I, Czajkowska J, Grzegorzek M (2015) Multi-scale textural feature extraction and particle swarm optimization based model selection for false positive reduction in mammography. Comput Med Imaging Graph 46:95–107

11. Roopa H, Asha T (2016) Segmentation of X-ray image using city block distance measure. In: 2016 International conference on control, instrumentation, communication and computational technologies (ICCICCT), Kumaracoil, pp 186–189
12. Roopa H, Asha T (2018) Feature extraction of chest X-ray images and analysis using PCA and kPCA. Int J Electr Comput Eng (IJECE) 8(5)
13. Haralick RM (1979) Statistical and structural approaches to texture. Proc IEEE 67(5):786–804

Salient Object Detection for Synthetic Dataset

Aashlesha Aswar and Arati Manjaramkar

Abstract Salient object detection essentially deals with various image processing and video saliency methodologies such as object recognition, object tracking, and saliency refinement. When image contains diverse object parts with cluttered background then using background prior we perform salient object detection through which we get more accurate and robust saliency maps. This paper introduces the analysis of salient object detection using synthetic dataset which also deals with negative interference of image that contains diverse object parts with cluttered background. Earlier study uses contrast prior but nowadays researchers use mainly boundary connectivity for improving the results. So, for detecting salient object we used four stages: first, we use SLIC superpixel method for image segmentation. Second, we use boundary connectivity which distinguishes the spatial layout of image region by considering image boundaries. Third, we use background measure and for reducing the noise in both foreground and background regions. Lastly, we use optimization framework through which we acquire a clean saliency map.

1 Introduction

Attention is the process of concentrating on one aspect of the environment and ignoring other things. So, visual attention has two stages: first the attention is distributed constantly over the external visual scene and processing of data is performed sequentially, second the attention is concentrated to a specific area of the visual scene and processing is performed serially. So, the word salient means most noticeable which is different from other parts at least in some aspect like color, texture, shape, etc. and object is nothing but a shape. There are various applications of visual attention according to researchers. So, the salient object detection is motivated by various

A. Aswar · A. Manjaramkar (✉)
Department of Information Technology, S.G.G.S. I E & T, Nanded, India
e-mail: akmanjaramkar@sggs.ac.in

A. Aswar
e-mail: aashleshaaswar@gmail.com

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_131

applications like automatic image cropping, image retargeting, and adaptive image display on small devices, image collection browsing, advertising design, image and video cropping, and object segmentation [1]. Recent studies witnessed that visual attention helps in object detection, tracking, and recognition [2].

Basically, all the methods use the properties of backgrounds and object, detect the object on the basis of contrast prior and background prior. Previously for detecting salient object almost all saliency methods uses contrast prior where the contrast between the object and their background region is high but recently background prior is used for saliency computation in which it is considered that the image boundary regions are preferably background [3].

While we using boundary prior we came across drawback because of that we consider all the image boundary is background and when the object is partially connected to the image boundary then it is not considered as salient object. For avoiding this drawback, we used boundary connectivity method. The detection procedure of salient object is categorized into four stages: first image segmentation, image boundary connectivity, salient object detection using background measure, and optimization.

Early method works on various datasets such as MSRA10K, ECSSD, DUT-OMRON, etc. Early models used human eye fixation datasets as researcher inspired by biological methodology.

Ma and Zhang [4] propose a framework which outputs three levels of attention such as attended view, areas, and points. They proposed a contrast-based saliency map and fuzzy growing algorithm which stimulates the selection of human perception, using this they can accurately locate attention areas. This model enlightens the borders so it is not used for detecting the salient object.

Basically, salient object detection algorithms main goal is to segment the salient object from image and it is evaluated using foreground mask and boundary boxes, these methods use low-level cues called contrast prior. Contrast prior uses the unique identity of the object and differentiate object and background using contrast between color, regions, etc.

Hemami et al. [5] detect the limitation and requirements of saliency maps for various previous methods. Based on the analysis they proposed a frequency-tuned approach which uses the low-level features such as color and luminance to compute saliency in images. Cheng et al. [6] propose histogram-based contrast (HC) which accomplished a faster result and high-quality saliency map is used for reducing cost of computational efficiency. Perazzi et al. [7] use a contrast and spatial distribution to obtain an accurate pixel saliency maps. They decompose image into homogeneous elements to compute saliency using Gaussian filter. Hence, the assumption is that the photographers do not crop the object from the image. So, the contrast prior is not useful. And, when the object boundary moderately touches to image boundary in that contest the boundary prior also fails or destroyed.

Wei et al. [9] propose the geodesic saliency (GS) which is used to get high accuracy and speed and which is also more robust to detect salient object. This method overcomes the limitations of background prior by connecting the image boundary regions using an edge with an appropriate boundary weight. Yang et al. [8] propose the manifold ranking on graph to detect the salient region and uses the two-stage approach with foreground and background measures or queries for ranking thorough which saliency map is generated.

Zhu et al. [9] propose a robust measure that uses both contrast prior and background prior. The minimization of optimization framework is used to get uniform salient object detection. And, the smoothness term is used to get clear object and background segmentation. Alexe et al. [10] measure the object characteristics in Bayesian framework and they propose an object measure which combines various measures. Zhang et al. [11] propose a cascaded architecture using the ranking SVM which generates an ordered set of proposals for windows that contains objects. Cheng et al. [12] propose BING, i.e., binarized version of gradient feature which generates the objectness proposals and it is tested using few atomic operations. Jiang et al. [13] propose a saliency detection algorithm which integrates three cues such as uniqueness, focusness, and objectness (UFO) and it gives top performance.

It is difficult to attribute the results to the specific algorithm properties because saliency maps are not smooth. Saliency filters show that using Gaussian filter they formulated contrast and saliency estimation in a combined way [7]. So, using boundary prior rather than obtaining background from image boundary, we focus on the object detection using objectness proposal techniques [12].

We then use a superpixel objectness measure and by thresholding this measure, we appropriately obtain background and foreground regions. Then, assign saliency values to superpixels using the proposed robust saliency measure called background connectivity. Then, we use an optimization framework rather than using weighted summation and multiplication or combining the cues. So, the optimization framework proposed by [9] in which the cost function assigns the values to superpixels and it is called foreground weights. And then, the minimization function is used to assign image background regions by taking lower values and assigning foreground regions with higher values. The smoothness function is used to uniform that obtained saliency maps in cost function.

We evaluated our method on various benchmark datasets: MSRA-1000 [5], SED1, SED2 [16], CSSD [2], and most importantly on synthetic images dataset. And proposed method also gives extraordinary results on synthetic image datasets.

Rest of the paper is organized as follows: Sect. 2 describes the methodology that contains segmentation and various methods to detect salient object. Section 3 describes the saliency optimization framework. Experimental results on various datasets are discussed in Sect. 4 and Sect. 5 concludes the paper.

2 Boundary Connectivity

2.1 Structuring Your Paper

Instead of assuming image patch as background or image boundary as background, we can easily connect the image boundary. The proposed method states that when the image patch region is deliberately connected to the image boundary then that image patch is background; otherwise, it is a salient object. This measure characterizes the spatial layout of image region with respect to image boundaries. It is stable in image content variations because it buses an instinctive geometrical interpretation and this also makes it unique.

The boundary connectivity measure detects the background at very high precision with very satisfactory recall using a single threshold. And, it also handles exclusively background images without objects and enhances a traditional contrast computation.

2.2 Definition

As compared to background regions, the object regions are much less connected to the image boundary that is why object and the background regions in general images are different in their spatial layout. This is shown in Fig. 1.

The synthetic image consists of four regions. As per the visual perception, the red region is clearly a salient object because it is big in size, compact, and only moderately connected to image boundaries. And, it is clear that the green and white regions are background because these regions are heavily connected to image boundaries. The pink region is too small in size and it is slightly connected to image boundary and that is why it is considered as a partially cropped object and it is not a salient object.

So, the boundary connectivity measure states that how heavily an image region R is connected to image boundary. Then, the boundary connectivity BondCon is as

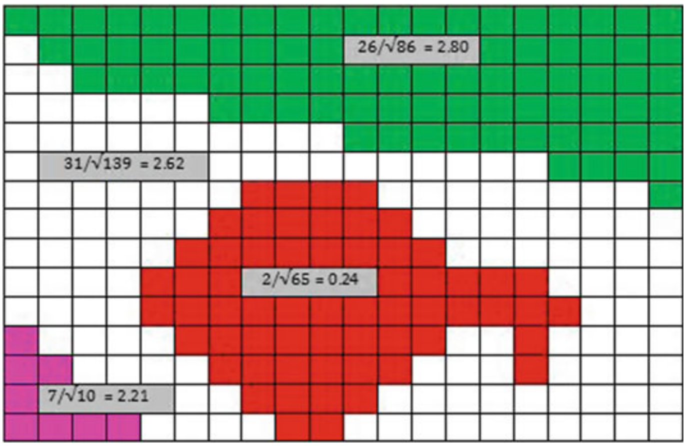


Fig. 1 Example of boundary connectivity. The synthetic image consists of four regions associated with BondCon values. BondCon is small for object region of image and more for background region of image

$$\text{Bondcon}(R) = \frac{|\{p|p \in R, p \in \text{Bond}\}|}{\sqrt{|\{p|p \in R\}|}} \tag{1}$$

where p is image patch, Bond is set of image boundary patches and it uses an instinctive geometrical interpretation. It is the ratio of regions perimeter on boundary to square root of its area or the regions overall perimeter. We used the square root of its area because the measure remains stable across different image patch resolutions. As shown in Fig. 1, the boundary connectivity is small for object regions and large for background regions.

2.3 Saliency Computation

Image segmentation is a process of portioning a digital image into multiple segments. The main goal of segmentation is to simply change the representation of an image into something which is more meaningful and easy to analyze. Figure 2 shows the flow of method, and Fig. 2b shows the output after applying SLIC superpixel segmentation.

Image segmentation itself is a challenging job; if we choose hard segmentation method, then it is very difficult for parameter and algorithm selection. So, we used a soft segmentation approach called SLIC method by [14] and it has high computational efficiency in which the image is abstracted as a set of superpixels. And, based on the experiments we find that 190–210 superpixels are sufficient for 300 * 400 resolution of image.

The SLIC superpixel method uses the CIELab color space. We connect all the adjacent superpixels (p, q) by constructing an undirected weighted graph and find the Euclidean distance between the average colors and it is denoted as $d_{\text{ap}}(p, q)$. Then, a collected weight with their shortest path on graph is called geodesic distance and is calculated between any two superpixels and is denoted as $d_{\text{ge}}(p, q)$.

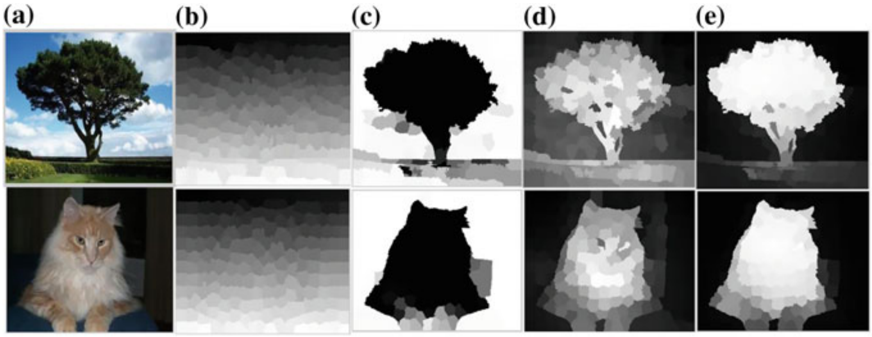


Fig. 2 Flow of method: **a** Input image, **b** SLIC superpixel segmentation, **c** Probability of background using Eq. (6), **d** Background weighted contrast using Eq. (7), **e** Optimized saliency map using Eq (8)

$$d_{ge}(p,q) = \min_{p_1 = p, p_2, \ldots, p_n = q} \sum_{i=1}^{n-1} d_{ap}(p_i, p_{i+1}) \tag{2}$$

If for comfort, we define $d_{geo}(p,q) = 0$. After this, we calculate the "spanning area" of each superpixel p as

$$\text{Area}(p) = \sum_{i=1}^{N} ex\left(-\frac{d_{ge}^2(p,p_i)}{2\sigma_{clr}^2}\right) = \sum_{i=1}^{N} S(p,p_i) \tag{3}$$

where N is the number of superpixels.

As per Eq. (3), the area (p) is computed as the soft area of the region where the image patch p belongs. We note that $S(p, p_i)$ in the summation is in (0,1) and distinguish how much superpixel p_i is subscribed to the p's area. We consider when superpixel p_i and p are in horizontal region, then note $d_{ge}(p,q) = 0$ and $S(p, p_i) = 1$ that ensures superpixel p_i adds a unit area to the area of p.

And when superpixel p_i and p are in different regions, then there is at least one edge that exists on their shortest path and $S(p, p_i) = 0$ that ensures superpixel p_i does not contribute to the area of p. So, by experiments, we find that the execution is strong when parameter σ is within (5–15) and then we set it to 12.

Correspondingly, we find the length along the boundary and set one for superpixels on image boundaries and set it to zero otherwise. And then, we compute the boundary connectivity of p as shown in Eq. (4).

$$\text{BondCon}(p) = \frac{\text{Len}_{bond}(p)}{\sqrt{\text{Area}(p)}} \tag{4}$$

where $\text{Len}_{bnd}(p)$ is the length of boundary for superpixel p. For calculating BondCon(p), the shortest path between all adjacent superpixel is calculated using Johnson's method [15].

Then, between any two superpixels add the edges. As shown in Fig. 3, the boundary connectivity values of background regions are enlarged [3]. This is useful due to closing of foreground objects and the physically connected background regions are separated.

Fig. 3 Image enhancement by connecting its boundaries: **a** Input image, **b** BondCon without linking image patches, **c** BondCon is improved by linking boundary patches

Then, the weighted contrast of background of image gives an important information for salient object detection and it enhances the general contrast computation. The regional contrast is calculated by the summation of contrast's overall appearance distance to all other regions to weight by its spatial distance [7].

$$\text{Contr}(p) = \sum_{i=1}^{N} d_{\text{ap}}(p, p_i) W_{sp}(p, p_i) \tag{5}$$

where $W_{\text{sp}}(p, p_i) = ex\left(-\frac{d_{\text{sp}}^2(p, p_i)}{2\sigma_{\text{sp}}^2}\right)$ is calculated using distance weighted matrix. The distance between two superpixels is denoted by $d_{\text{sp}}^2(p, p_i)$. Figure 2c shows the result after applying Eq. (5).

Then, we present a new term called probability of background. It is calculated by mapping the superpixels boundary connectivity value and it is shown in Eq. (6).

$$W_i^{\text{bkg}} = 1 - ex\left(-\frac{\text{BondCon}^2(p_i)}{2_{\text{bondCon}}^2}\right) \tag{6}$$

When boundary connectivity value is very large, its value is near to one, otherwise, its value is near to zero. And the probability of background is denoted by W_i^{bkg}.

So, for better result, we set BondCon to be one but if we set the value of BondCon in between 0.5 and 2.5, then our result is inconsiderable because it would not affect the previous result. Then, we calculate the summation of distance appearance with spatial distance and probability of weight to enhance the contrast and it is shown in Eq. (7).

$$w\text{Contr}(p) = \sum_{i=1}^{N} d_{\text{ap}}(p, p_i) W_{\text{sp}}(p, p_i) W_i^{\text{bkg}} \tag{7}$$

As per Eq. (7) object region of image will get large probability of background from background region of image and then their contrast increases, and on the other side the background region of image will get very less probability of background from the object region of an image and then their contrast is reduced. Figure 2d shows the result after applying Eq. (7).

3 Saliency Optimization

In general, there are various saliency cues which are merged using assigning weight to summation and multiplication. Alternately, we use a framework to optimize saliency. In saliency optimization, we merge the background weight and foreground weight used in [7]. And, it is also merged with smoothness function and it is stated as

$$\sum_{i=1}^{N} W_i^{\mathrm{fg}}(t_i - 1)^2 + \sum_{i=1}^{N} W_i^{\mathrm{bkg}}(t_i)^2 + \sum_{i,j} W_{ij}(t_i - t_j)^2 \tag{8}$$

where Eq. (8) shows that the ultimate value of saliency which is assigned to p_i is t_i and after the minimization of smoothness cost function W_i^{bkg} is denoted as the background weight corresponding to superpixel p_i and W_i^{fg} is denoted as the foreground weight corresponding to superpixel p_i. High foreground weight W_i^{fg} supports superpixel p_i to take values near to one and high W_i^{bkg} supports to take the value of p_i near to zero. And, W_{ij} is a smoothness term for which we used the same parameter setting which is used in [3]. Figure 2e shows the result after applying Eq. (8).

4 Experimental Results

We evaluated the proposed approach on four benchmark datasets and a synthetic image dataset. These datasets are MSRA-1000 [5], SED1, SED2 [16], where MSRA-1000 contains large images with complex background and this is the most widely used dataset for salient object detection. SED1 is a dataset which contains the images with only one object and on the other hand, the SED2 dataset which contains the images with two objects and both datasets contain the object with different sizes and datasets. Synthetic images datasets contain 40 images which contain 3D objects icons, etc. Results on all datasets are compared ground-truth mask labeled by humans. Figure 4 shows experimental results of proposed methods on different datasets.

4.1 Precision and Recall

We evaluate a method using precision–recall curves and MAE, i.e., mean absolute error. Precision is the fraction of relevant instances means pixels assigned correctly among the retrieved instances means against the total number of pixels assigned salient, whereas recall is the fraction of relevant instances that is retrieved as the pixels labeled correctly in relation over total relevant instances in the image means to the number of ground-truth pixels.

Table 1 shows that using threshold as two, the proposed measure can detect background with very decent recall and high precision. Figures 5 and 6 show the performance of boundary connectivity and geodesic saliency based on precision and recall for synthetic datasets.

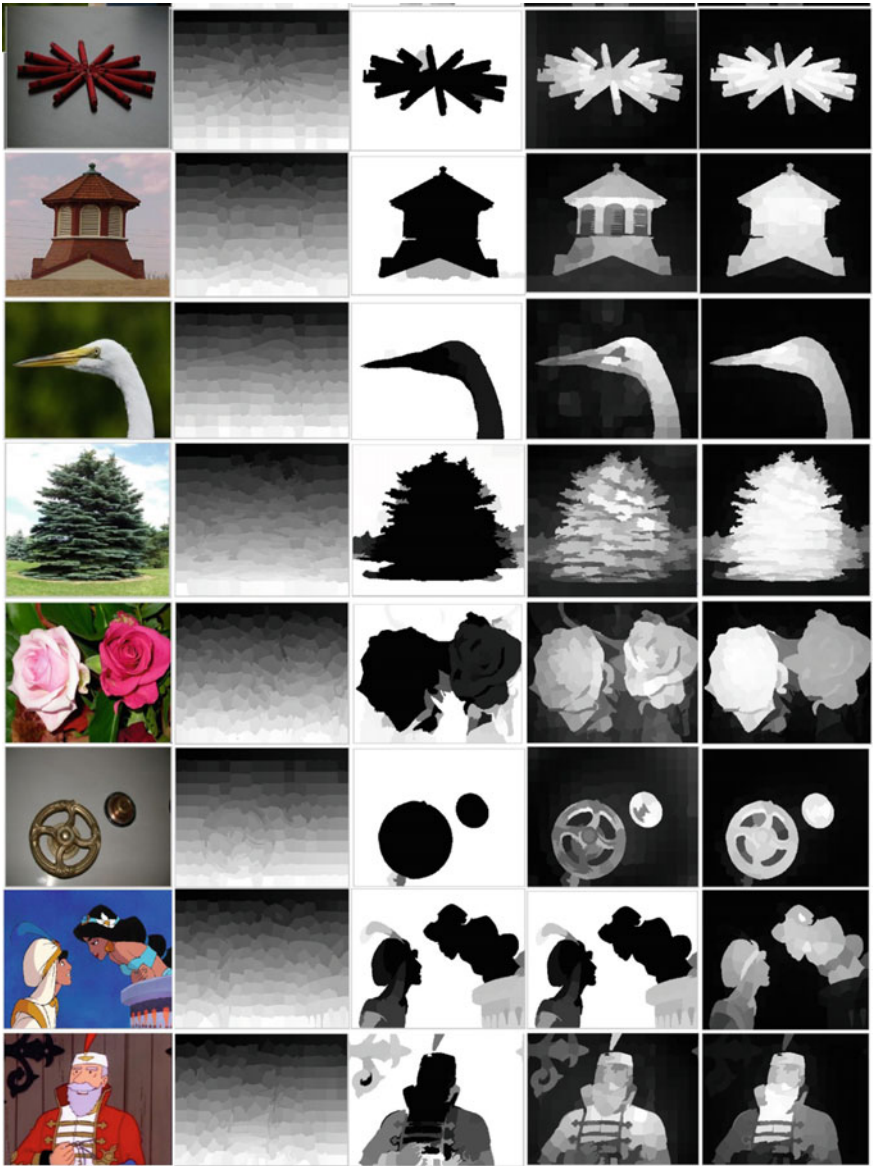


Fig. 4 Experimental results of proposed methods on different datasets

Table 1 Background precision/recall for superpixels with boundary connectivity>2 on three benchmark datasets

Benchmark	Boundary connectivity		Geodesic saliency	
	Precision (%)	Recall (%)	Precision (%)	Recall (%)
MSRA	98.3	77.3	98.3	63.6
SED1	97.4	81.4	96.5	69.6
SED2	95.8	88.4	94.7	65.7

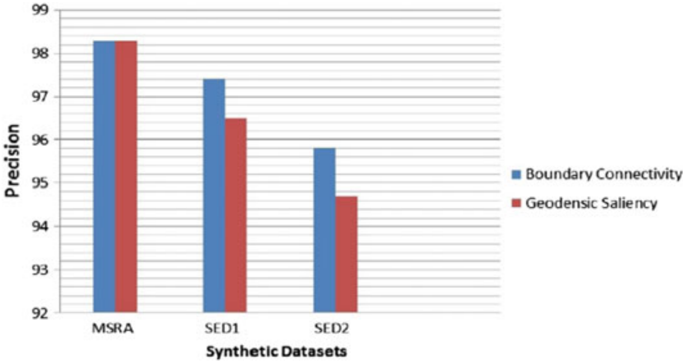


Fig. 5 Performance of boundary connectivity and geodesic saliency based on precision for synthetic datasets

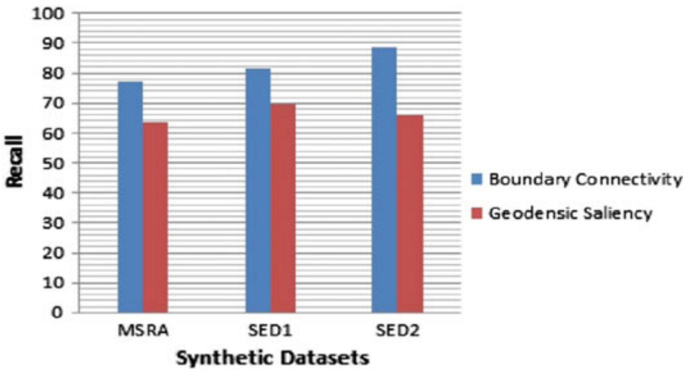


Fig. 6 Performance of boundary connectivity and geodesic saliency based on recall for synthetic datasets

4.2 Mean Absolute Error

Mean absolute error (MAE) measures the similarity between saliency map and the ground truth of images [7]. For a saliency map S, ground-truth mask is G, then MAE is defined as

$$\mathrm{MAE} = \frac{1}{w * h} \sum_{x=1}^{h} \sum_{y=1}^{w} |S(x, y) - G(x, y)| \tag{9}$$

where h and w denote the height and weight of image. The proposed approach performs better than the state-of-the-art methods as per MAE.

5 Conclusion

In this paper, we used a background measure to detect salient object which produces very clean and precise saliency maps. Using boundary connectivity, the background region is produced instinctively. Using SLIC superpixel segmentation it is very easy to segment the image. Optimization method is used for reducing the noise from image and smoothing it. We perform experiments on various benchmark datasets such as MSRA [5], SED1, SED2 [16], synthetic datasets and achieve state-of-the-art results. In future work, we develop some advanced method to utilize background prior.

References

1. Liu T, Sun J, Zheng N, Tang X, Shum H (2007) Learning to detect a salient object. In: CVPR
2. Yan Q, Xu L, Shi J, Jia J (2013) Hierarchical saliency detection. In: CVPR
3. Zhu W, Ling S, Wei Y, Sun J (2014) Saliency optimization from robust background detection. In: CVPR
4. Ma Y-F, Zhang H-J (2003) Contrast-based image attention analysis by using fuzzy growing
5. Achanta R, Hemami S, Estrada F, Susstrunk S (2009) Frequency-tuned salient region detection, pp 1597–1604
6. Cheng M-M, Zhang G-X, Mitra NJ, Huang X, Hu S-M (2011) Global contrast based salient region detection, pp 409–416
7. Perazzi F, Krähenbühl P, Pritch Y, Hornung A (2012) Saliency filters: contrast based filtering for salient region detection, pp 733–740
8. Lu H, Ruan X, Yang C, Zhang L, Hsuan Yang M (2013) Saliency detection via graph-based manifold ranking
9. Zhu W, Wei Y, Wen F, Sun J (2012) Geodesic saliency using background priors
10. Alexe B, Deselaers T, Ferrari V (2012) Measuring the objectness of image windows 34(11)
11. Zhang Z, Warrell J, Torr PHS (2011) Proposal generation for object detection using cascaded ranking svms, pp 1497–1504
12. Cheng M-M, Zhang Z, Lin W-Y, Torr PHS (2014) BING: Binarized normed gradients for objectness estimation at 300 fps. In: IEEE CVPR
13. Jiang P, Ling H, Yu J, Peng J (2013) Salient region detection by ufo: Uniqueness, focusness and objectness. In: ICCV'13, pp 1976–1983
14. Machiras V, Decenciere E, Walter T (2015) Spatial repulsion between markers improves watershed performance. Mathematical morphology and its applications to signal and image processing. Springer International Publishing, pp 194–202
15. Johnson DB (1977) Efficient algorithms for shortest paths in sparse networks. J ACM 24(1):1–13
16. Alpert S, Galun M, Basri R, Brandt A (2007) Image segmentation by probabilistic bottom-up aggregation and cue integration. In: CVPR

Atherosclerotic Plaque Detection Using Intravascular Ultrasound (IVUS) Images

A. Hari Priya and R. Vanithamani

Abstract The atherosclerotic plaque deposition in artery is a type of cardiovascular disease and is a major factor of death. Mostly larger and high-pressure vessels such as the femoral, cerebral, renal, coronary, and carotid arteries are influenced by atherosclerosis. Hence, the characterization of plaque distribution and its liability to rupture are mandatory to judge the degree of risk and to schedule the treatment. Intravascular ultrasound (IVUS) is an ultrasound imaging modality which uses a unique catheter. The catheter will be provided with an ultrasound probe which is miniaturized and is connected to the lateral end. In this paper, segmentation of the IVUS image using fast marching method (FMM) is done and will be followed by the feature extraction. Feature extraction methods such as local binary pattern (LBP), speeded up robust feature (SURF), and histogram of oriented gradients (HOG) are used and the resulting image is classified using Euclidean distance classifier. By comparing the results of subjected feature extraction techniques, LBP method is found suitable for the detection of atherosclerotic plaque.

1 Introduction

Atherosclerotic plaque formation is a diseased condition of an artery which is characterized by the narrowing of the lumen in initial stages followed by the building up of plaques. During the initial stages, there will be no symptoms recorded. A fatty substance develops on the surface of the inner walls of arteries which is known as plaque. It causes thickening of the artery and the artery may lose its elasticity in

A. Hari Priya (✉)
Medical Electronics, Avinashilingam Institute for Home Science
and Higher Education for Women, Coimbatore, Tamil Nadu, India
e-mail: haripriyaarivukkarasu@gmail.com

R. Vanithamani
Biomedical Instrumentation Engineering, Avinashilingam Institute
for Home Science and Higher Education for Women, Coimbatore, Tamil Nadu, India
e-mail: vaniraj123@yahoo.co.in

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_132

further stages. Further plaque deposition will lead to the narrowing of artery. It may also result in thrombosis. The artery may be blocked suddenly or embolus may be formed out of this thrombosis. This embolus (condition in which the clot travels along with the blood) may lead to stroke. Partial block of the artery by plaque may lead to increased blood pressure in that artery [1].

Many of the risk factors for atherosclerosis can be controlled and factors such as age and family history cannot be changed. Smoking, devoid of physical exercise, unhealthy diet, and over consumption of alcohol can be controlled. Another major cause for atherosclerosis is the obesity [2].

IVUS is a catheter-based imaging system that allows the technicians or healthcare professionals or interventional cardiologists to get the images of diseased blood vessels from lumen of the vessel [3]. Lumen characteristics, size of the vessel, diseased artery segments, and characteristics of plaques can be obtained from IVUS images [4, 5].

In this work, different feature extraction methods such as local binary pattern (LBP), HOG, and SURF were used to detect the atherosclerotic plaque using an IVUS image. Then, the suitable feature extraction technique to detect plaque is identified [6].

2 Materials and Methods

Processing of the IVUS image is done to detect the plaque formation inside an artery. It is a step-by-step process which includes the segmentation of IVUS image, followed by feature extraction and classification. The flow diagram of the IVUS image processing for identification of abnormal artery is given in Fig. 1

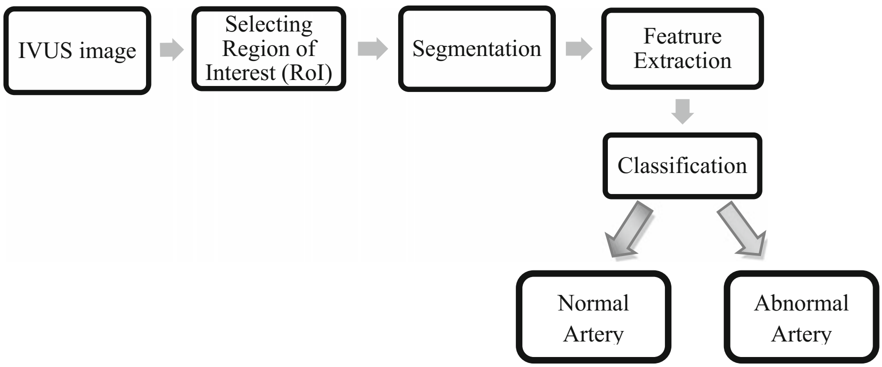


Fig. 1 Flow diagram of the IVUS image processing

2.1 Region of Interest

Input image is chosen from the dataset and then the RoI is selected manually. To exclude the lumen from the processing, the RoI is selected. In future, lumen can be automatically removed from the consideration to increase the quality of processing.

2.2 Segmentation

Partitioning the digital image into divergent regions is termed as image segmentation. The segmentation process will turn the input image into an image with more useful and meaningful information. Segmentation can locate the edges and boundaries more typically [7]. Image segmentation allocates a label to each pixel of the image and pixel with the same label is grouped which may share some similar characteristics. Fast marching method (FMM) is used to segment the IVUS image [8].

FMM It solves the boundary value problems numerically [8]. The eikonal equation is given in Eqs. 1 and 2

$$|\nabla u(x)| = 1/f(x) \quad \text{for}(x \text{ belongs to } \Omega) \tag{1}$$

$$U(x) = 0\,\text{for}(x \text{ belongs to } \delta\Omega) \tag{2}$$

For non-flat domain problems, Eq. 3 is used

$$|\Delta su(x)| = 1/f(x) \tag{3}$$

FMM algorithm The domain is partitioned into mesh and the mesh points are referred as nodes. Each node is assigned as x_i and the corresponding value is represented in Eq. 4

$$U_i = U(x_i) \approx u(x_i) \tag{4}$$

Nodes are named as far (nodes not visited yet), considered (nodes that are visited and tentatively assigned), and accepted (nodes which are visited and values are permanently assigned).

Step 1 For every node x_i, the value of U_i is given in Eq. 5

$$U_i = +\infty \tag{5}$$

Label these x_i nodes as far and for nodes $x_i \in \delta\Omega$ set $U_i = 0$ which is labeled as accepted.

Step 2 For all nodes x_i use the eikonal update formula to compute the new value for $\widehat{U}$. If $\widehat{U} < \widehat{U}_i$, then set $U_i < \widehat{U}$ and label x_i as considered.
Step 3 Assume $\dot{x}$ be the node considered with the lowest value U. $\dot{x}$ is labeled as accepted.
Step 4 For every neighbor x_i of $\dot{x}$ which is not accepted, the tentative value $\widehat{U}$ is calculated.
Step 5 If $\widehat{U} < \widehat{U}_i$ then set U_i as $\widehat{U}$. If x_i was labeled as "Far", then the label should be considered as updated.
Step 6 If there is a "Considered" node, return to step 3 or abort.

2.3 Feature Extraction

Feature refining or extraction starts from an initial set of deliberate data and builds features that are considered to be revealing and necessary which can lead to better manual explications [9]. Grouping a subgroup of the embryonic features is termed as feature selection. The determined features should have the pertinent information from the input data. Thus, the desired process can be done using diminished representation instead of using the whole data for further processing [10, 11]. In this paper, feature extraction techniques like LBP, HOG, and SURF are used to refine features from the segmented IVUS image.

LBP LBP is a feature refining technique which uses the neighboring pixels to extract features. It is considered to be a substantial feature refining technique for texture classification [12, 13].

LBP algorithm.

Step 1 Divide the segmented image into pixels.
Step 2 Select a pixel and compare the nominated pixel to all neighboring pixels (eight neighboring pixels). It is done for all pixels starting from a corner pixel and follows the pixel along clockwise or counterclockwise circle.
Step 4 The particular bit in the binary array is assigned either to 0 or to 1. If the considered pixel's value is more than the neighboring pixel's value, then assign "0" or assign "1".
Step 5 The binary number thus obtained is converted into the decimal number. This resultant decimal number is stored in the center pixel. Thus, LBP mask is created by processing each pixel possessed by the input image.
Step 6 LBP histogram is calculated where the LBP mask value ranges from 0 to 255. And so, the obtained LBP descriptor will be of 1×256 in size.
Step 7 Concatenate the histogram values to get the features extracted.

Histogram of Oriented Gradient (HOG) HOG is used to recognize the objects. This method of feature extraction counts the incidence of gradient orientation value in every cell of the image. As it is employed on local cells, the HOG is advantageous over the other descriptors. As it acts on a local cell, HOG is stable to photometric and geometric transformations [13].

HOG Algorithm The HOG algorithm involves five steps such as

Step 1 Compute the gradient values. A point distinct derivative mask is applied in one or both of the directions (horizontal and vertical) to get the gradient values. The filter kernels used are $[-1,0,1]$ and $[-1,0,1]^T$.

Step 2 In this step, the cell histograms are created. With the values got from the gradient calculation, each pixel (within the cell) launches a weighted value for a histogram channel (orientation based).

Step3 Locally, normalization of the gradient strengths is done to record changes in lighting and contrast. The normalization of gradient strength is processed by grouping the cell into spatially connected blocks.

Step 4 Let the non-normalized vector be v, possessing all histograms (within a given block), $\|v\|k$ be the k-norm for $k = 1, 2$ and e be a small constant. Equation 6 is used as normalizing factor.

$$f = \frac{v}{\sqrt{\|v\|_2^2 + e^2}} \tag{6}$$

Step 5 The HOG can be adapted for recognition of an object by providing them as features.

Speeded Up Robust Features (SURF) SURF is used to observe and recognize objects, registration of image, and classification of data. It employs Hessian blob detector and Haar wavelet response to refine features from the given input data. Hessian blob detector is used to observe the interest points. The total of Haar wavelet result around the considered point is the feature descriptor of the SURF method of feature extraction [14].

SURF Algorithm The surf algorithm involves the following steps:

Step 1 Square-shaped filters are used as Gaussian smoothing approximation.

Step 2 The Hessian blob detection is done. The Hessian matrix determinant is the measure of local change around the considered point. The maximal determinant points are chosen. For a point $p = (x, y)$ in an image A, the Hessian matrix $H(p, \sigma)$ is

$$H(p, \sigma) = \begin{pmatrix} Lxx(p, \sigma) & Lxy(p, \sigma) \\ Lyx(p, \sigma) & Lyy(p, \sigma) \end{pmatrix} \tag{7}$$

where the elements of the matrix are second-order derivative's convolution of Gaussian with the image $A(x, y)$ at the x point.

Step 3 Images are smoothened repeatedly using Gaussian filter and then they are subsampled to reach the higher level of the pyramid, thus consecutive stairs with various mask measures are calculated.

$$\sigma \text{ approx} = \text{current filter size} \times (\text{base filter scale}/\text{base filter size})$$

Step 4 The reproducible orientation has been fixed based on data from a region of circle around the considered point. Then, a region of square is assigned to the orientation selected and the SURF descriptors are refined from it.

Step 5 The Haar wavelet results within a circular locality in both x-direction and y-direction of radius 6 s near (around) the considered point are calculated where s is the scale with which the considered point is detected. Sum the vertical and horizontal responses within the particular window. The local orientation vector is calculated by summing the two responses. The placement of the considered point is defined by the longest of the vectors calculated.

Step 6 The interest region is further divided into smaller square regions, then the Haar wavelet results are refined at sample points which are regularly spaced. To gain more robustness, the resultant riposte is weighted with Gaussian.

2.4 Classification

Classification is the technique of gathering the testing data into a preset class based on the trained data set containing observations whose class is already known [15]. In this paper, Euclidean distance measurement is used for classification.

Euclidean Distance The straight-line space between two pixels is said to be the Euclidean distance. The space between two points in Euclidean space is the Euclidean distance [16]. In 2-D space, the distance between $R = (a_1, a_2)$ and $S = (b_1, b_2)$ is defined as

$$\sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2} \tag{8}$$

3 Results and Discussion

The suggested method is tested using the IVUS image obtained from the database available at https://www.dropbox.com/s/el674ocdp9uojro/Training_Set.zip?dl=0.

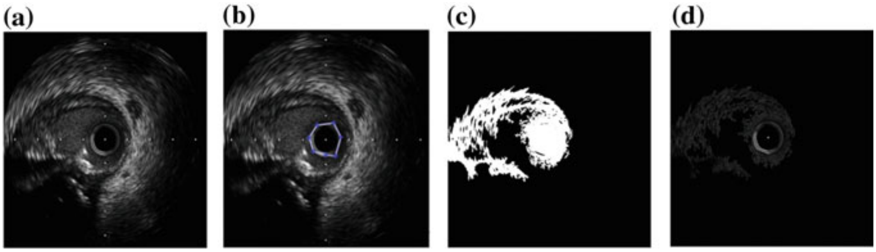


Fig. 2 Segmentation of test image-1 using FMM method: **a** Original image, **b** Selection of RoI, **c** Inverted image, **d** Segmented image

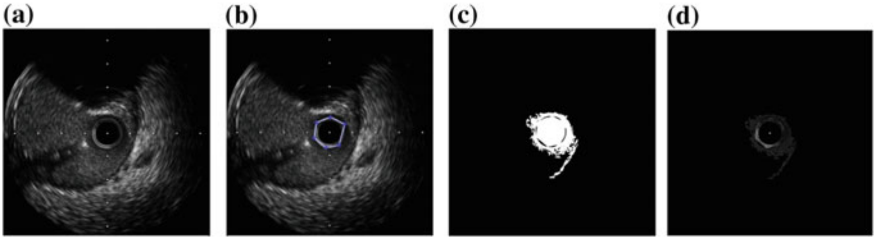


Fig. 3 Segmentation of test image-2 using FMM method: **a** Original image, **b** Selection of RoI, **c** Inverted image, **d** Segmented image

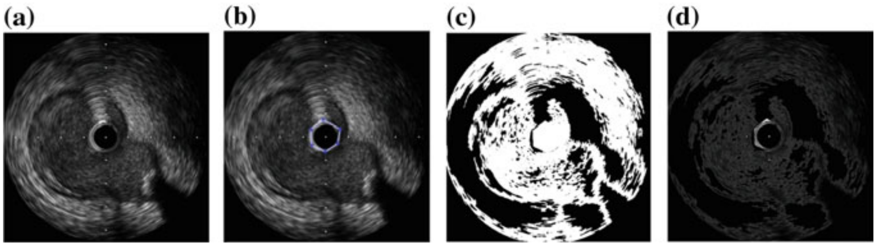


Fig. 4 Segmentation of test image-3 using FMM method: **a** Original image, **b** Selection of RoI, **c** Inverted image, **d** Segmented image

MATLAB is used to test the algorithm. Twenty-five images are taken to validate the results. The segmented results of test images 1, 2, 3, and 4 are shown in the figures from Figs. 2, 3, 4, and 5.

The results are tabulated in Table 1.

From the results obtained, the sensitivity can be found using Eq. 9

$$\mathrm{TP}/(\mathrm{TP} = \mathrm{FN}) \tag{9}$$

The specificity can be found using Eq. 10

$$\mathrm{TN}/(\mathrm{TN} + \mathrm{FP}) \tag{10}$$

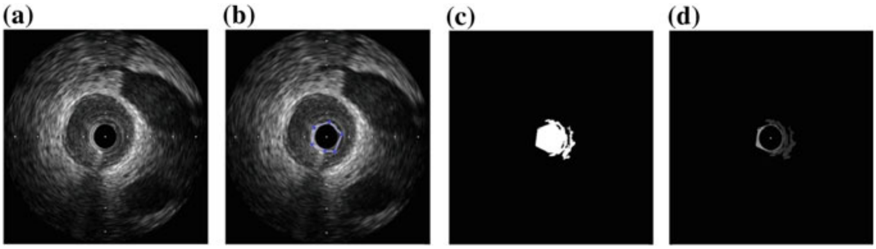


Fig. 5 Segmentation of test image-4 using FMM method: **a** Original image, **b** Selection of RoI, **c** Inverted image, **d** Segmented image

Table 1 Results of the tested artery using different feature extraction techniques

Algorithm	Classification results			
	TP	TN	FP	FN
LBP	7	1	1	16
HOG	1	7	1	16
SURF	3	5	1	15

Table 2 Sensitivity, specificity, and accuracy of detection of abnormal artery using LBP, HOG, and SURF techniques

Algorithm	Sensitivity	Specificity	Accuracy (%)
LBP	0.3043	0.5	92
HOG	0.0589	0.875	68
SURF	0.1666	0.8333	72

Equation 11 can be used to find the accuracy

$$TP = FN/(TP + FP + TN + FN) \quad (11)$$

The sensitivity, specificity, and accuracy for detecting the abnormal artery using LBP, HOG, and SURF are given in Table 2.

3.1 Conclusion

Atherosclerotic plaque can make arteries narrower, leading to reduced or blocked blood flow which in turn may cause heart failure [17]. Thus, it is very mandatory to diagnose the atherosclerotic arteries accurately [18]. In this work, the IVUS images are segmented using fast marching segmentation. Three feature extraction techniques such as LBP, HOG, and SURF are used. From the results obtained, the LBP is found to be a more precise method for detecting atherosclerotic artery compared to HOG and SURF. Thus, the LBP is a better technique to process the IVUS image than the HOG and SURF techniques for detecting atherosclerotic plaque.

References

1. Taki A, Roodaki A, Setarehdan SK, Avansari S, Unal G, Navab N (2013) An IVUS image-based approach for improvement of coronary plaque characterization. Comput Biol Med 43:268–280
2. Taki A, Roodaki A, Pauly O, Setarehdan SK, Unal G, Navab N (2009) A new method for characterization of coronary plaque composition via IVUS IMAGES. IEEE-2009, pp 787–790
3. Zhang Q, Wang Y, Ma J, Shi J (2011) Contour detection of atherosclerotic plaques in IVUS images using ellipse template matching and particle swarm optimization. In: 33rd annual international conference of the IEEE EMBS, pp 5174–5177
4. Chen F, Ma R, Liu J, Zhu M, Liao H (2018) Lumen and media-adventitia border detection in IVUS images using texture enhanced deformable model. Comput Med Imaging Graph 66:1–13
5. Zakeri FS, Setarehdan SK, Norouzi S (2017) Automatic media-adventitia IVUS image segmentation based on sparse representation framework and dynamic directional active contour model. Comput Biol Med 89:561–572
6. Shi Y, Witte RS, O'Donnell M (2005) Identification of vulnerable atherosclerotic plaque using IVUS-based thermal strain imaging. IEEE Trans Ultrason Ferroelectr Freq Control 52:844–850
7. Balocco S (2014) Standardized evaluation methodology and reference database for evaluating IVUS image segmentation. Comput Med Imaging Graph 38:70–90
8. Wien TU, Ives St (2016) Comparison of the parallel fast marching method, the fast iterative method, and the parallel semi-ordered fast iterative method. In: The international conference on computational science, vol 80, pp 2271–2275
9. Krishnan S, Athavale Y (2018) Trends in biomedical signal feature extraction. Biomed Signal Process Control 43:41–63
10. Anam S, Misawa H, Uchino E, Suetake N (2012) Parameter tuning by PSO for fuzzy inference-based coronary plaque extraction in IVUS image. IEEE-2012 2012:1426–1429
11. Roodaki A, Taki A, Setarehdan SK, Navab N (2008) Modified wavelet transform features for characterizing different plaque types in IVUS images. A feasibility study. IEEE-2008, pp 789–792
12. Mahale VH, Ali MH, Yannawar PL, Gaikwad AT (2017) Image inconsistency detection using local binary pattern (LBP). In: 7th international conference on advances in computing & communications, vol 115, pp 501–508
13. Korkmaz SA, Binol H (2018) Classification of molecular structure images by using ANN, RF, LBP, HOG, and size reduction methods for early stomach cancer detection. J Mol Struct 1156:255–263
14. Kan S-C, Cen Y-G, Cen Y, Wang Y-H, Voronin V, Mladenovic V, Zeng M (2017) SURF binarization and fast codebook construction for image retrieval. J Vis Commun Image Represent 49:104–114
15. Caballero KL, Barajas J, Pujol O, Rodriguez O, Radeva P (2007) Using reconstructed IVUS images for coronary plaque classification. In: Proceedings of the 29th annual international conference of the IEEE EMBS, pp 2167–2170
16. Deza MM, Deza E, Marie M (2009) Encyclopedia of distances. Springer, pp 94
17. Li X, Li J, Jing J, Ma T, Liang S, Zhang J, Mohar D, Raney A, Mahon S, Brenner M, Patel P, Kirk Shung K, Zhou Q, Chen Z (2014) Integrated IVUS-OCT imaging for atherosclerotic plaque characterization. IEEE J Sel Top Quantum Electron 20
18. Dehnavi SM, Babu MSP, Yazchi M, Basij M (2013) Automatic soft and hard plaque detection in IVUS images: a textural approach. In: Proceedings of 2013 IEEE conference on information and communication technologies, pp 214–219

Correlative Feature Selection for Multimodal Medical Image Fusion Through QWT

J. Krishna Chaithanya, G. A. E. Satish Kumar and T. Ramasri

Abstract A novel image fusion technique is proposed in this paper to achieve an efficient and informative image by combining multiple medical images into one. This method accomplishes quaternion wavelet transform (QWT) as a feature representation technique and correlation metric for feature selection. Here, the correlation-based feature selection is accomplished to extract the optimal feature set from the sub-bands thereby to reduce the computational time taken for fusion process. QWT decomposes the source images first and then the feature selection process obtains an optimal feature set from the low-frequency (LF) as well as high-frequency (HF) sub-bands. The obtained optimal feature sets of both LF sub-bands and HF sub-bands are fused through low-frequency fusion rule and high-frequency fusion rule and the fused LF and HF coefficients are processed through IQWT to obtain a fused image. Various models of medical image are processed in the simulation and the performance is evaluated through various performance metrics and compared with conventional approaches to show the robustness and efficiency of proposed fusion framework.

J. Krishna Chaithanya (✉) · G. A. E. S. Kumar
Department of ECE, Vardhaman College of Engineering, Hyderabad, India
e-mail: j.krishnachaitanya@vardhman.org

T. Ramasri
Department of ECE, SVUCE, SVU, Tirupati, India

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_133

1 Introduction

Medical imaging has attained a vast research interest in recent years due to the increasing requirement of disease diagnosis and clinical investigation. Owing to various medical image modalities, every image modal can illustrate some prior information about the human body, but limited to specific purpose only. For instance, the structures of bones and hard tissues are better visualized through the computed tomography (CT) image, whereas the detailed structure of soft tissues is illustrated through magnetic resonance imaging (MRI) only. In the same manner, the details of anatomical structures are represented through MRI-T1 image model, whereas the normal tissues and pathological tissues are described through MRI-T2 only [1–4]. Further, image models including the functional magnetic resonance imaging (fMRI), positron emission tomography (PET), and single photon emission computed tomography (SPECT) provide the details about the functional information in low-resolution spatial images, useful in the detection of metabolic abnormalities and cancer-related diseases. Analysis of individual imaging model builds an extra manual burden and also results in an increased time complexity. Hence, there is a necessity to combine the multiple image models to construct a compendious figure. Medical image fusion is a compendious technique which gives a fused image as an input by fusing multiple images with different modalities. On the other hand, the fused images are more suitable to help the doctor in the treatment planning: fusion of CT and MRI images can represent the bone structures and soft tissues in order to represent the physiological and anatomical features of a human body simultaneously. A further advantage with image fusion is the reduction in the storage cost.

In past, numerous techniques are developed to achieve an efficient fused image which provides complete information about different multimodal images. The entire earlier image fusion techniques are divided into transform domain techniques and spatial domain techniques. In the case of spatial domain approaches, the fused image directly relates the pixel intensities of source image, whereas in the transform domain approaches, these are related with their transformed coefficients. Since there exists a direct relation among the pixels of input and output image, the small variations in the source image intensities result in a proportional effect in the fused image. For instance, the common drawback observed form the earlier spatial domain approaches is the reduced contrast in the fused image which represents the bad quality fused image. Due to this, the portions which need to be highlighted in the fused image would not get highlighted and the regions which need to be suppressed will get highlighted. Instead of directly operating over the pixel intensities, the transform-domain-based fusion approaches perform over transformed pixel intensities.

A novel image fusion technique is proposed in this paper to achieve an efficient and informative image by combining multiple medical images into one. This method accomplishes quaternion wavelet transform (QWT) as a feature representation technique and correlation for feature selection. The QWT is more robust to phase variations and the proposed correlation-based feature selection is robust for computational time. The source image is decomposed into sub-bands through QWT first, and then the obtained bands are processed for correlation-based feature selection and then the selected coefficients are fused into their respective modified sub-bands. Finally, the modified sub-bands are formulated as a fused mage through IQWT.

Reminder of the paper is formulated as follows: The related literature survey of the proposed work is illustrated in Sect. 2. The basic preliminaries about QWT and correlation are described in Sect. 3. The details of the proposed image fusion method are illustrated in Sect. 4. Section 5 illustrates the details of performance evaluation, and finally Sect. 6 concludes the paper.

2 Literature Survey

2.1 Spatial Domain Approaches

The approaches proposed based on principal component analysis (PCA), intensity–hue–saturation (IHS), and Brovey transform are better examples for spatial domain approaches [5–7]. To deal with sparse representation of medical image, Zhang et al. [6] proposed an image fusion technique for medical images centered to PCA and some sparse representation techniques named as SPCA. In this approach, the correlations between and within the channels are modeled effectively. Here, the sparse representation is applied over remote sensing image and the PCA is applied to achieve a reduced dimensionality to reduce the computational cost. Kaur et al. [8] proposed a PCA-based image fusion technique in the collaboration with genetic algorithm. However, all these techniques fail in the spectral degradation. Another region segmentation approach is proposed in [9] to find the regions through the morphological filtering. According to [9], the source images are initially processed for fusion through the simple averaging method. Then, the obtained fused image is segmented into different regions through the normalized cut method. Further, based on the obtained regions, the source images are subjected to region segmentation and then the segmented regions are processed for final fusion according to their spatial frequencies with respect to the regions of initial fused image. Though this method is observed to be too simple, it introduces unnecessary side effects in the fused image like reduced contrast level [10].

2.2 Transform Domain Approaches

Recently, the transform base image fusion approaches have gained a lot of research interest due to its effectiveness in the quality increment of fused images. Among those transforms, multiscale transforms (MST) is one of the most popular transforms which transforms the image in the multiscale fashion. Framelet transform [11], discrete wavelet transform (DWT) [3, 12, 13], Gabor transform, non-subsampled contourlet transform (NSCT) [1, 4, 14], and contourlet transform [15] are some example of MSTs. Compared to DWT, the CT represents the details of image features more precisely by which the performance of fusion approach increases. A further edge perspective fusion framework is proposed by Deng et al. [16] based on the canny edge operator and DWT. This approach mainly aims to preserve the edge features. DWT decomposes the source image initially into the low-frequency and high-frequency (vertical, horizontal, and diagonal) sub-bands. Further, the canny operator is applied over vertical and diagonal sub-band images to acquire the edge information. An effective DWT-based multi-focus image fusion approach is proposed by Yang et al. [17] by seeing the physical meaning of wavelets. A novel coefficient selection algorithm is also described in this approach to select the optimal feature set for fusion. Zhou et al. [18] presented a novel multiscale fusion method based on weighted gradient to solve the image misregistration problem which results in the degradation of multi-focus images. Further, to solve the misregistration problems at the fused images, this approach focused on the two-scale mechanism, i.e., focusing on both large scale and small scale coefficients. Considering the limitations in the directionality of 2D separable wavelets, the WT cannot represent the edges directions more accurately.

3 Preliminaries

3.1 Quaternion Wavelet Transform (QWT)

Considering the analytics of quaternion signal [19], axioms of the quaternion algebra and the separability property, the QWT is derived as a natural extension from the real and complex wavelet transforms. The QWT is a perfect transform technique for the signals with higher dimensions. In the case of QWT, the signal with larger dimensions can be decomposed into multiresolution levels by direct accomplishment. Since the QWT results in phase-directed bands along with amplitude band, a more detailed analysis can be acquired over the image with different phase variations. This is limited in the conventional wavelets as well as in the complex wavelets. Thus, instead of a distance similarity accomplishment in the complex wavelet pyramid, QWT uses phase concept for top-down parameter estimation.

The QWT of an image can be defined as

$$f(x, y) = A_n^q f(x, y) + \sum_{s=1}^{n} \left[D_{s,1}^q f(x, y) + D_{s,2}^q(x, y) + D_{s,3}^q f(x, y)\right] \quad (1)$$

where A_n^q and $D_{s,p}^q, (p = 1, 2, 3)$ are the approximation sub-band images and detailed sub-band images. The analytical extension of QWT is based on the analytical expression of real wavelet and its two-dimensional Hilbert transform, as represented by

$$\begin{cases} \psi^D(x,y) = \psi_h(x)\psi_h(y) \\ \rightarrow \psi^D + iH_{i1}\psi^D + jH_{i2}\psi^D + kH_{i3}\psi^D \\ \psi^V(x,y) = \phi_h(x)\psi_h(y) \\ \rightarrow \psi^V + iH_{i1}\psi^V + jH_{i2}\psi^V + kH_{i3}\psi^V \\ \psi^H(x,y) = \psi_h(x)\phi_h(y) \\ \rightarrow \psi^H + iH_{i1}\psi^H + jH_{i2}\psi^H + kH_{i3}\psi^H \end{cases} \tag{2}$$

$$\begin{cases} \phi(x,y) = \phi_h(x)\phi_h(y) \\ \rightarrow \phi + iH_{i1}\phi + jH_{i2}\phi + kH_{i3}\phi \end{cases} \tag{3}$$

whereϕ is scaling function, ψ^D, ψ^V, ψ^H are the wavelet functions oriented to diagonal, vertical and horizontal directions, respectively. From 2D Hilbert transform, 1D Hilbert transform can be derived through the x- and y-axis as

$$\begin{cases} \psi_h, \psi_g = H\psi_h \\ \phi_h, \phi_g = H\phi_h \end{cases} \tag{4}$$

Based on the above analytical extension, the 2D QWT can be defined as

$$\begin{cases} \psi^D(x,y) = \psi_h(x)\psi_h(y) + i\psi_h(x)\psi_g(y) \\ \quad +j\psi_g(x)\psi_h(y) + k\psi_g(x)\psi_g(y) \\ \psi^V(x,y) = \phi_h(x)\psi_h(y) + i\phi_g(x)\psi_h(y) \\ \quad +j\psi_g(x)\phi_h(y) + k\phi_g(x)\psi_g(y) \\ \psi^H(x,y) = \phi_h(x)\psi_h(y) + i\phi_g(x)\psi_h(y) \\ \quad +j\psi_g(x)\phi_h(y) + k\phi_g(x)\psi_g(y) \\ \phi(x,y) = \phi_h(x)\phi_h(y) + i\phi_g(x)\phi_h(y) \\ \quad +j\phi_g(x)\phi_h(y) + k\phi_g(x)\phi_g(y) \end{cases} \tag{5}$$

where the representations given in first three rows of above QWT equation [20, 21] are the mathematical formulae to extract the high-frequency coefficients from the image along diagonal, vertical, and horizontal direction, respectively. Further, the last row is for the extraction of QWT low-frequency coefficients from the image.

3.2 Correlation Measure

Once both low approximation sub-bands and detailed sub-bands are obtained through QWT, they are needed to be fused to obtain a single approximation and detailed sub-band image. One important point to be noticed is that the obtained bands consist of most redundant information, especially in the high-frequency bands. The QWT extract totally four bands, one is amplitude and three are phase-oriented bands. Further, all the three phase bands are oriented in three directions such as horizontal, vertical, and diagonal at the detailed band extraction. If all these bands are processed for fusion process, it results in a more computational burden over the system and also consumes too much time to obtain a fused image. Hence, this work proposed a new feature selective coding based on the correlations of features in all the bands.

Further, the fusion needs to be applied over the coefficients which are more informative and also should not affect its visual quality. To achieve a better fusion performance, this work accomplished directional filters by which the variations in all directions are predicted much efficiently. Once the regions with fewer variations are predicted through directional filters, those regions are processed for fusion. For this purpose, the correlation is measured between all the coefficients of all the sub-bands. The coefficients with maximum correlation are only selected for fusion. This is applied over every sub-band and only a set of feature is extracted from every band. In the case of proposed QWT-based fusion framework, the QWT decomposes the image into low frequencies and high frequencies followed by magnitude and three phase bands. The proposed correlation measure evaluation is applied over every band and a new feature set with maximum correlation is extracted for both source images A and B, then the extracted new and optimal feature set of both images are processed for fusion. The simple architecture of correlative-measure-based feature selection is shown in Fig. 1.

For this purpose, the proposed correlative-measure-based coefficient selection algorithm is developed and termed as correlative feature selection (CFS). In this way, this approach of coefficient selection results in the feature set with optimal nature and is also highly informative with less burden.

4 Fusion Framework

The following step-by-step process illustrates the detailed explanation of the proposed image fusion technique. Initially, consider the input images with two different models A and B. Apply the proposed fusion mechanism over these two images to obtain a fused image as follows:

Step 1 Apply QWT over both source images A and B to decompose them into the low-frequency (Approximations, A) and high-frequency (Details, D) sub-band images. Apply orientational filter over both LF and HF sub-band images to extract the wavelet coefficients in particular orientation *l*. Let us

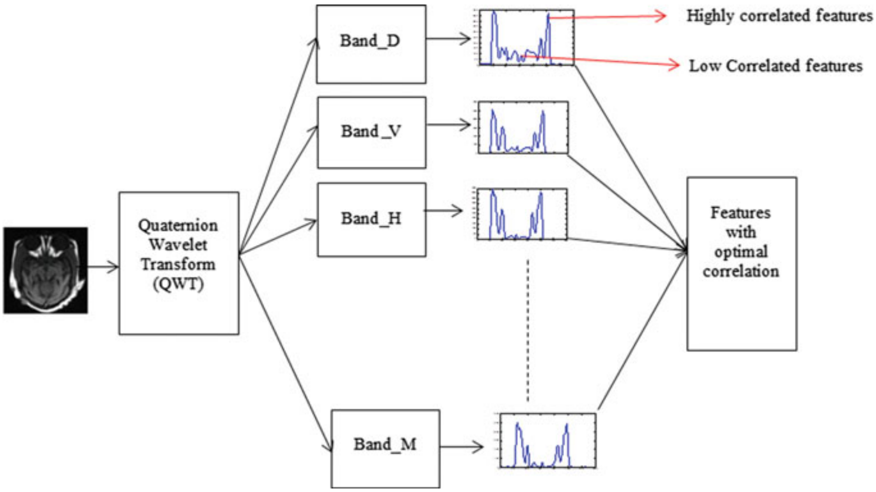


Fig. 1 Schematic of correlation-based feature selection

consider the approximate sub-band $A^{A}_{k,l}$ which is of the input image A and the approximate sub-band $A^{B}_{k,l}$ which is of the input image B at kth orientation and lth direction, respectively. Further, consider the detailed sub-band image $D^{A}(x, y)$ is of the input image A and $D^{B}(x, y)$ is of the input image B. Further, apply the CFS over all the obtained sub-band images.

Step 2 Perform the CFS over the approximation bands of both images to obtain an optimal feature set by which the extra complexity over the system reduces.

Step 3 Perform fusion of approximation sub-band images $A^{A}_{k,l}$ and $A^{B}_{k,l}$ through the proposed fusion rule based on the angular consistency [22] to obtain a fused approximation sub-band image, $A^{F}(x, y)$.

In general, the approximate sub-bands fusion is carried out through the most popular "averaging rule." However, due to an issue of reduced contrast level in the fused image, this rule is violated and a novel fusion rule is derived considering the angular consistencies of approximation sub-bands. The proposed angular consistency rule for the fusion of approximation sub-bands is formulated as

$$A^{F}(x, y) = \begin{cases} A^{A}(x, y) & \text{if } A^{A\theta}_{c}(x, y) > A^{B\theta}_{c}(x, y) \\ A^{B}(x, y) & \text{if } A^{A\theta}_{c}(x, y) < A^{B\theta}_{c}(x, y) \\ \frac{1}{2}(A^{A}(x, y) + A^{B}(x, y)) & \text{if } A^{A\theta}_{c}(x, y) = A^{B\theta}_{c}(x, y) \end{cases} \quad (6)$$

where $A^{A\theta}_{c}(x, y)$, $A^{B\theta}_{c}(x, y)$ are the angular consistencies of approximations of input images A and B, correspondingly.

Step 4 Perform the CFS over the approximation bands of both images to obtain an optimal feature set by which the extra complexity over the system reduces.

Step 5 Perform fusion of detailed sub-band images $D^A(x, y)$ and $D^B(x, y)$ based on new fusion rule proposed through the spatiofrequency energies of detailed sub-bands [23] to obtain a fused detailed or high-frequency sub-band image, $D^F(x, y)$.

In general, the detailed sub-band image fusion is carried out through the most popular "Larger Absolute Selection Rule." However, due to the loss of vast complementary information in the fused image, this rule is not used here and the new fusion is derived based on the spatiofrequency energies of detailed sub-bands. Particularly, in the case of detailed sub-bands, preservation of edges, lines are very much important which gives a more detailed analysis about the soft tissues in medical images. Further, discrimination between noise and detailed coefficients is more important by which the fusion process may mislead to consider a noise as a high-frequency coefficient and considers it as a fusing coefficient. Hence, there is a need to provide a perfect discrimination between noises and high-frequency coefficients, the proposed spatiofrequency energies take this responsibility. The proposed spatiofrequency energies-based fusion rule is formulated as

$$D^F(x, y) = \begin{cases} D^A(x, y) \text{ if } E^A(x, y) \geq E^B(x, y) \\ D^B(x, y) \text{ if } E^A(x, y) < E^B(x, y) \end{cases} \quad (7)$$

where $E^A(x, y)$ and $E^B(x, y)$ represent the "Spatiofrequency energies" of detailed sub-bands of input images A and B, respectively.

Step 6 Recreate the fused image by applying inverse QWT over the fused approximation sub-band $A^F(x, y)$ and detailed sub-band $D^F(x, y)$.

5 Simulation Results

This section illustrates the details of performance evaluation of the proposed framework quantitatively and qualitatively. To verify the performance of the proposed approach, an extensive simulation is carried out over various types of images like medical images, natural images, etc. Here, the source images of size 256 * 256 are considered and the simulation is carried through MATLAB software. The obtained fused images are shown in Figs. 2, 3, 4, 5, and 6.

Sources of the test images are as follows:

1. The Whole Brain Atlas (Harvard): http://www.med.harvard.edu/aanlib/
2. BrainWeb: Simulated Brain Database: http://brainweb.bic.mni.mcgill.ca/brainweb/

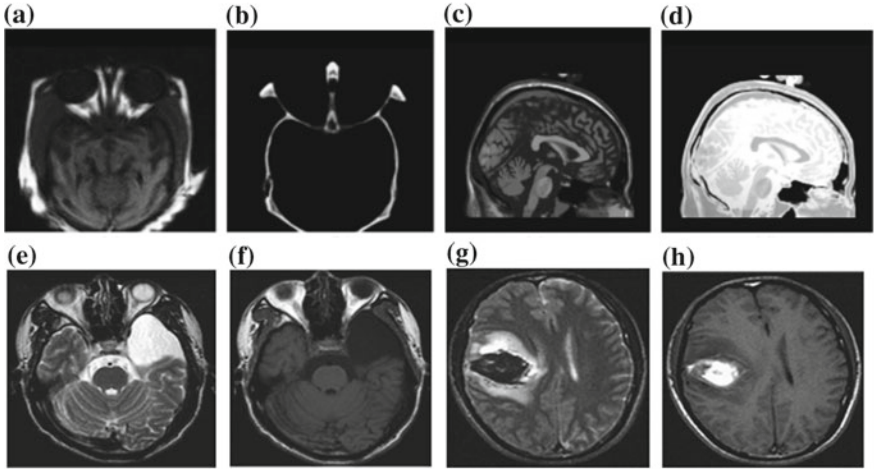


Fig. 2 Test images: **a**, **c** Magnetic resonance imaging (MRI), **b**, **d** Computed tomography (CT) images, **e**, **g** Magnetic resonance-T2 relaxation (MR-T2) images, and **f**, **h** Magnetic resonance-T1 (MR-T1) relaxation images

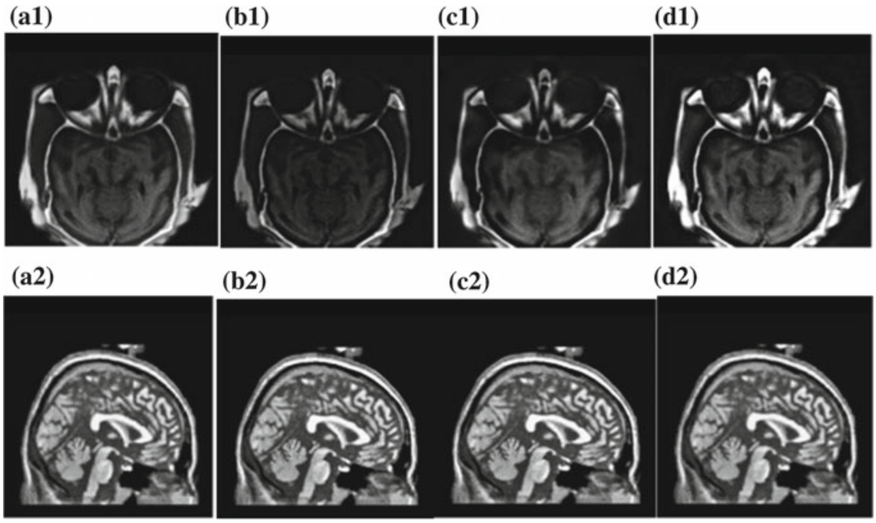


Fig. 3 Obtained results for CT and MRI images: **a**1–**a**2 Fused images through DWT, **b**1–**b**2 Fused images through CT, **c**1–**c**2 Fused images through NSCT, and **d**1–**d**2 Fused images through proposed

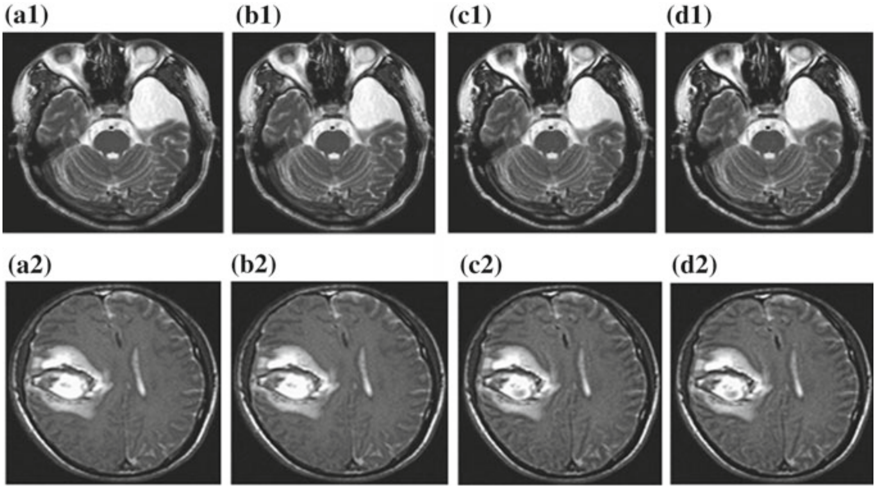


Fig. 4 Obtained results for MR-T1 and MR-T2 images: **a**1–**a**2 Fused images through DWT, **b**1–**b**2 Fused images through CT, **c**1–**c**2 Fused images through NSCT, and **d**1–**d**2 Fused images through proposed

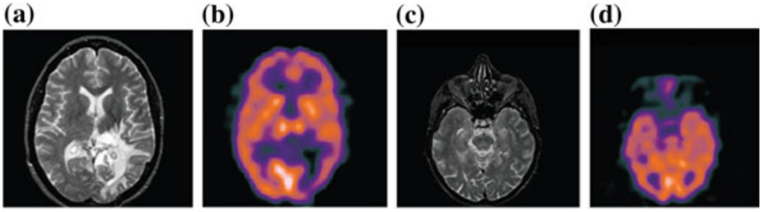


Fig. 5 Source images: **a**, **c** MRI images, **b**, **d** SPECT images

Under the performance evaluation, five objective evaluation measurement parameters are adopted to evaluate the fusion performance. There are weighted quality fusion index (WQFI) [24], local quality index (LQI) [24], edge-dependent fusion quality index (EFQI) [24], and $_{/}$ [25] which measures the transmission of edge features and visual features from input images to fused images, and mutual information (MI) [26] which measures the amount of information transferred from input images to output fused images. The range ofQ_0, Q_w, Q_E, and $Q_{AB/F}$ lies between 0 and 1 and the range of MI is above 1 and it varies from image to image. These performance metrics are evaluated for all the above test image sets and are formulated in Figs. 7, 8, 9, 10, and 11.

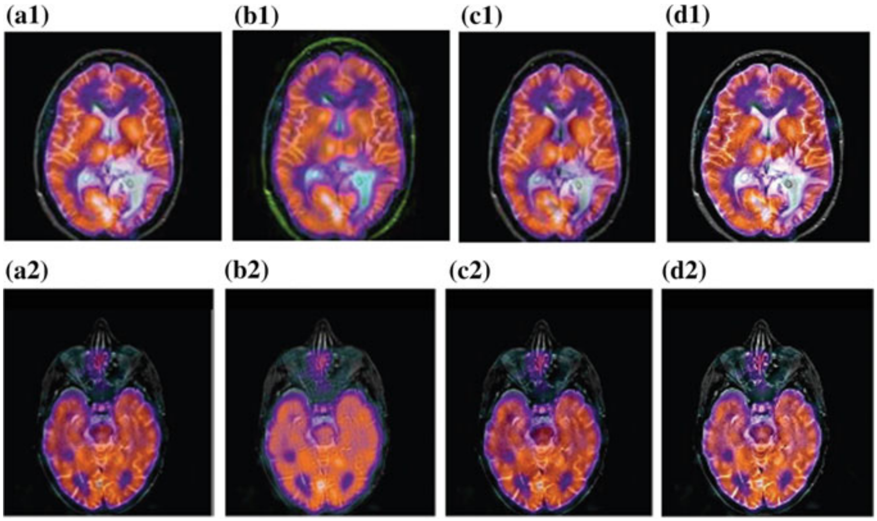


Fig. 6 Obtained results for MR-T1 and MR-T2 images: **a**1–**a**2 Fused images through DWT, **b**1–**b**2 Fused images through CT, **c**1–**c**2 Fused images through NSCT, and **d**1–**d**2 Fused images through proposed

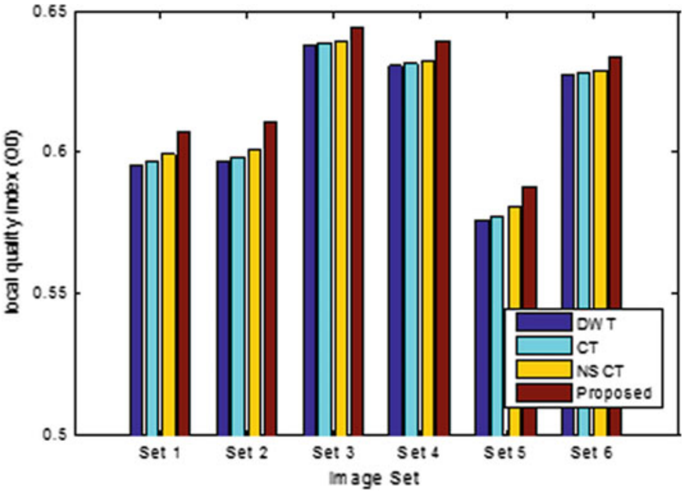


Fig. 7 Local quality index comparison

6 Conclusion

A novel image fusion technique is proposed in this paper to achieve an efficient and informative image by combining multiple medical images into one. This method accomplishes QWT as a feature representation technique and correlation for feature

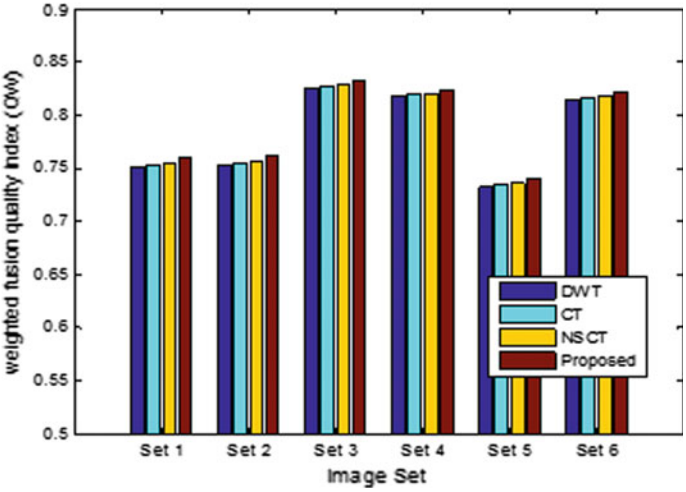


Fig. 8 Weighted fusion quality index comparison

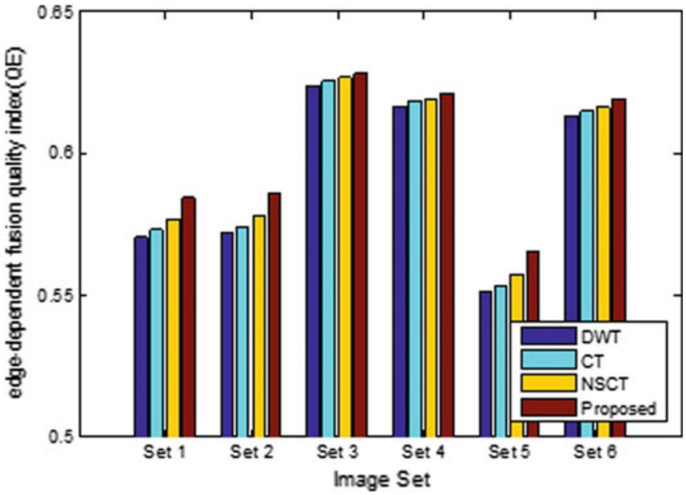


Fig. 9 Edge-dependent fusion quality index

selection. Compared to the conventional multiscale transform techniques such as DWT, CT, and NSCT, the QWT decomposes the image into the phase-deviated bands also by which more information will be revealed about the frequency characteristics of image. Further, the proposed correlation-based feature selection optimizes the feature set to be fused, reduces the extra computational time. The obtained simulation result shows an outstanding performance.

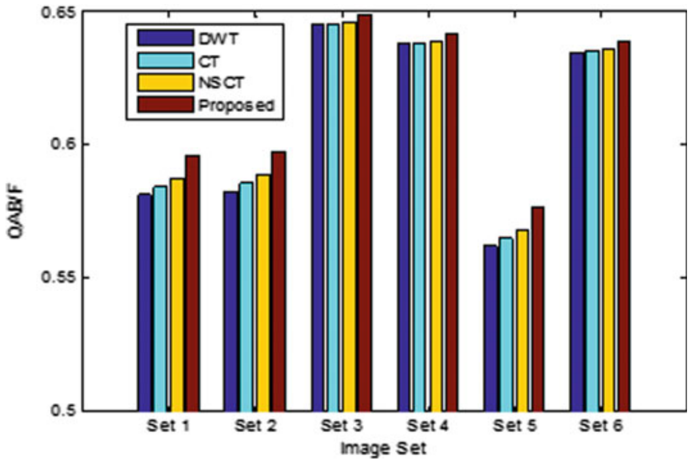


Fig. 10 QAB/F comparison

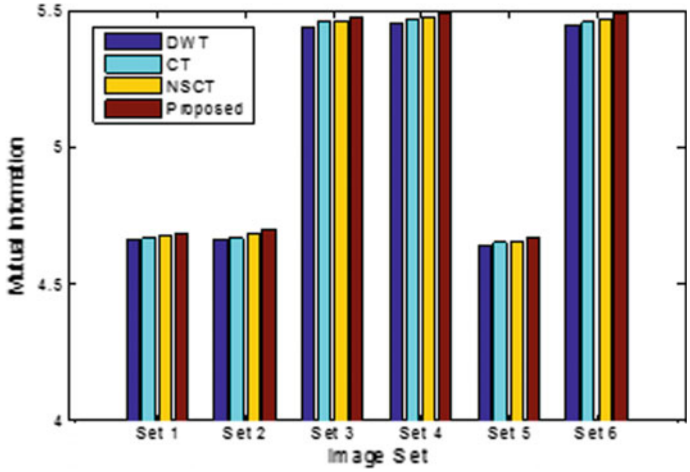


Fig. 11 Mutual information comparison

References

1. Gomathi PS, Kalaavathi B (2016) Multimodal medical image fusion in non-subsampled contourlet transform domain. Circ Syst 7(8):1598–1610
2. Valdés Hernández MDC, Ferguson KJ, Chappell FM, Wardlaw JM (2010) New multispectral MRI data fusion technique for white matter lesion segmentation: method and comparison with thresholding in FLAIR images. Eur Radiol 20(7):1684–1691
3. Liu Z, Yin H, Chai Y, Yang SX (2014) A novel approach for multimodal medical image fusion. Expert Syst Appl 41(16):7425–7435
4. Yang Y, Tong S, Huang S, Lin P (2014) Log-Gabor energy based multimodal medical image fusion in NSCT domain. Comput Math Methods Med 2014:1–12

5. Saravanan V, Babu G, Sivakumar R, Monie EC (2013) Medical image fusion by PCA method and implementation on FPGA. Int J Emerg Trends Sci Technol 03(05):800–806
6. Zhang X (2016) Sparse representation and PCA method for image fusion in remote sensing. In: 2016 2nd international conference on control, automation and robotics (ICCAR)
7. Salem YB (2016) Image fusion models and techniques at pixel level. In: Image processing, applications and systems (IPAS)
8. Kaur R (2016) An approach for image fusion using PCA and genetic algorithm. Int J Comput Appl 145(6):(0975–8887)
9. Li S, Kang X, Hu J, Yang B (2013) Image matting for fusion of multi-focus images in dynamic scenes. Inf Fus 14(2):147–162
10. Liu Y, Jin J, Wang Q, Shen Y, Dong X (2014) Region level based multi-focus image fusion using quaternion wavelet and normalized cut. Sig Process 97:9–30
11. Bhatnagar G, Wu QMJ, Liu Z (2013) Human visual system inspired multi-modal medical image fusion framework. Expert Syst Appl 40(5):1708–1720
12. Shen R, Cheng I, Basu A (2013) Cross-scale coefficient selection for volumetric medical image fusion. IEEE Trans Biomed Eng 60(4):1069–1079
13. Li S, Yang B (2008) Multi-focus image fusion using region segmentation and spatial frequency. Image Vis Comput 26:971–979
14. Ganasala P, Kumar V (2014) CT and MR image fusion scheme in non-subsampled contourlet transform domain. J Digit Imaging 27(3):407–418
15. Yang L, Guo BL, Ni W (2008) Multimodality medical image fusion based on multiscale geometric analysis of contourlet transform. Neurocomputing 72(1–3):203–211
16. Deng Ai (2011) An image fusion algorithm based on DWT and canny operator. Adv Res Comput Educ Simul Model 175:32–38
17. Yang Y, Huang SY, Gao J, Qian Z (2014) Multi-focus image fusion using an effective discrete wavelet transform based algorithm. Meas Sci Rev 14(2):102–108
18. Zhou Z, Li S, Wang B (2014) Multi-scale weighted gradient-based fusion for multi-focus images. Inf Fusion 20:60–67
19. Chan WL, Choi H, Baraniuk R (2004) Quaternion wavelets for image analysis and processing. In: Proceedings of international conference image processing, pp 3057–3060
20. Yin M, Liu W, Shui J, Wu J (2012) Quaternion wavelet analysis and application in image denoising. Math Prob Eng 2012, Art. no. 493976
21. Gai S, Wang L, Yang G, Yang P (2016) Sparse representation based on vector extension of reduced quaternion matrix for multiscale image denoising. IET Image Process 10(8):598–607
22. Yang Y, Tong S, Huang S, Lin P (2014) Log-Gabor energy based multimodal medical image fusion in NSCT domain. Comput Math Methods Med 2014:1–12
23. Yang Y, Tong S, Huang S, Lin P (2015) Multi-focus image fusion based on NSCT and focused area detection. IEEE Sens J 15(5):2824–2838
24. Piella G, Heijmans H (2003) A new quality metric for image fusion. In: Proceedings of the international conference on image processing (ICIP '03), Catalonia, Spain, pp 173–176, September 2003
25. Xydeas CS, Petrović V (2000) Objective image fusion performance measure. Electron Lett 36(4):308–309
26. Qu G, Zhang D, Yan P (2002) Information measure for performance of image fusion. Electron Lett 38(7):313–315

Early Detection of Proliferative Diabetic Retinopathy in Neovascularization at the Disc by Observing Retinal Vascular Structure

Nilanjana Dutta Roy and Arindam Biswas

Abstract Proliferative Diabetic Retinopathy (PDR) is the advanced stage of Diabetic Retinopathy (DR) with high risk of severe visual impairment. Neovascularization is a common scenario at this stage where abnormal vessels proliferate. This paper describes a semi-automated method for early detection of PDR around few diameters of Optic Disc (OD) in retinal images. Center of OD detection from segmented images is essentially important here because the approach focuses on Neovascularization at the Disc (NVD). Around OD center on few pixel distance window boundary, the width of major vessels are measured and counted. Finally, the major vessels are identified by distinct colors. The sensitivity and specificity results on STARE dataset of 25 images are 0.86 and 0.87, respectively. The approach shows the average accuracy as 0.88.

1 Introduction

Diabetic Retinopathy (DR) is one of the leading causes of blindness for patients who are suffering from diabetes for many years. Out of four major progressive stages of DR, Proliferative Diabetic Retinopathy (PDR) is the advanced stage where lack of oxygen supply in blood vessels triggers the proliferation of thin and fragile blood vessels. These new fragile vessels, which are most likely to leak and bleed, when they grow around certain diameter of Optic Disc (OD) region in human eye are classified as Neovascularization at the Disc (NVD). When they grow at anywhere else in the vasculature, are termed as neovascularization elsewhere (NVE). Both the categories are responsible for abnormal proliferation and can cause vitreous hemorrhage,

N. D. Roy (✉)
Department of Computer Science and Engineering,
Institute of Engineering and Management, Kolkata, India
e-mail: nilanjanaduttaroy@gmail.com

A. Biswas
Department of Information Technology,
Indian Institute of Engineering Science and Technology, Shibpur, Howrah, India
e-mail: barindam@gmail.com

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_134

which further leads to a high risk of vision loss. Although a lot of algorithms have been developed to detect PDR in retinal images, the number is significantly less in comparison to non Proliferative Diabetic Retinopathy (NPDR) detection. A recent approach in [1] shows the extraction of vessel patterns and OD region for NVD/NVE classification by multilayered threshold. This method is dependent on a few structural and intensity based features for classification. Monitoring the openness of the major temporal arcade was proposed by the authors of [2] using single and dual parabolic models in normal and PDR images. To classify the stages of neovascularization, morphological transformations and Gaussian filtering were used in [3]. But all these methods described above, are prone to showing high false positive rates due to lack of analysis on the classification of features for NVD/NVE separately. The detection of NPDR lesions was successfully done using Hessian-based filtering approaches proposed by the authors of [4], but the value of such filters for PDR detection has not been significantly analyzed yet.

In this work, we present preprocessing and segmentation, followed by the center of OD detection, count and measurement of the width of major vessels, and coloring them in the Methodology section. Validation stage is to monitor the images regularly. Experimental evaluation section deals with the results, performance measurement, and databases we have used. An overall conclusion is stated in the Conclusion section.

This paper makes a contribution on detecting early signs of NVD PDR based on retinal vessel structure. This could further help the ophthalmologists to screen the disease and start early treatment.

An image of normal eye and image of PDR effected eye where the presence of fragile vessels are clearly seen around OD are shown in Fig. 1.

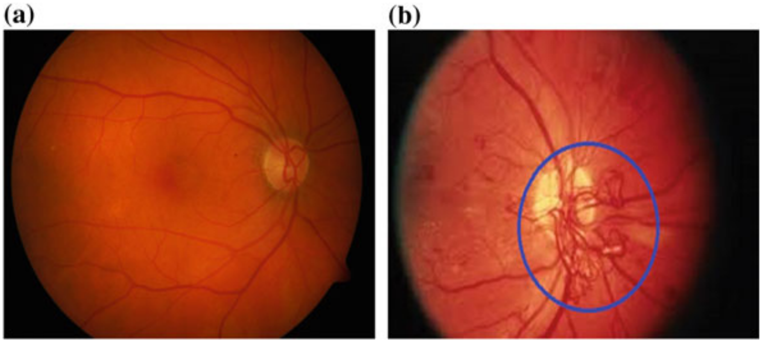


Fig. 1 Fundus images **a** normal eye, **b** PDR effected eye with NVD

2 Methodology

The task of PDR detection is accomplished based on the segmentation process followed by OD center detection, count on major vessels around OD, width measurement, and vessel coloring techniques. Initial steps include preprocessing to enhance the vasculature of the image and to ensure the image's quality against noise. Thereafter, the need for segmentation comes to validate the presence of abnormality in the image. Regular analysis of the eye of a diabetic patient, within few diameter areas around OD is indicative of the unusual changes about to happen. As the purpose of this approach is to detect neovascularization at OD (NVD), we are using the center of OD as the prime reference point from where all the vessels emanate. Repetitive count on major vessels around OD and measuring the width of them also help as a significant indication toward proliferation of fragile vessels. Conflict in vessel count raises an issue of abnormal situation whereas latter help in catching newly created tiny and fragile vessels. Coloring each vessel is used as an additional feature for validation applied over the resultant image along with the local features. The framework produces a final decision by comparing changes in number, width, and color of the tiny vessels with the stored image of the same person in the database.

2.1 Image Preprocessing and Segmentation

The fundus images from DRIVE [5] database, captured by Canon CR5 non-mydriatic 3 CCD cameras with a 45° FOV for medical imaging, have to undergo few image operations. Some common image processing steps are applied here to make the image ready for further processing. Grayscale conversion, sharpening using multiple passes of illumination distribution by CLAHE, followed by Otsu thresholding helped us to give the image a good shape at its initial stage. The grayscaled image is then passed through 2D median filtering for de-noising and finally, a smooth textured image in the binary platform is hence presented. Figure 2 shows the process for segmentation.

2.2 Center of Optic Disc

Regular monitoring on OD and its neighborhood region quicken the process of initial PDR detection as NVD originates at the optic disc region. So to track the progress, the center of OD detection plays an important role here as a reference point. In a 3×3 sized scanning window over the segmented binary image, a white pixel is detected as terminal point whose more than seven neighbors are black. Starting from the terminal points, all the straight lines are removed from the vasculature by making them as black till the junction points arrive. The resultant image is a structural 'ring' with all the closed polygons and with no single straight line. Later, the junction points are

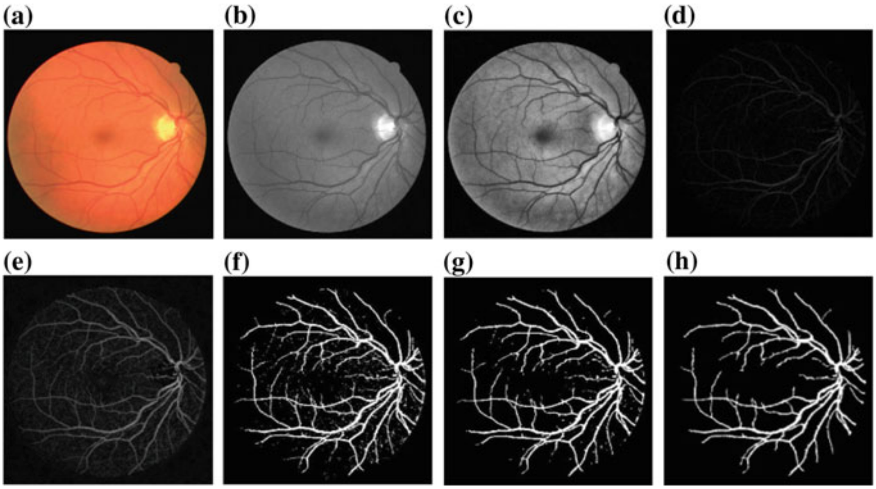


Fig. 2 **a** Original RGB image, **b** green channeled image, **c** AHE filtered image, **d** image after bottom hat, **e** contrast enhanced image, **f** extracted blood vessels by Otsu's thresholding, **g** median filtered image, **h** segmented image after noise removal

detected from the "ring". It is observed that the number of junction points increase at around few diameters of OD. Here, a cluster is identified with maximum number of junction points and a convex hull [6] is formed out of it. The centroid of the resultant convex hull, shown in Eqs. 1 and 2, is detected as the reference center of OD. A flow of this section is shown in Fig. 4. A result of center of optic disc detection is shown in Fig. 3a.

$$C_x = \frac{1}{6A}\sum_{i=1}^{N-1}(x_i + x_{i+1})(x_i y_{i+1} - x_{i+1} y_i) \tag{1}$$

and

$$C_y = \frac{1}{6A}\sum_{i=1}^{N-1}(y_i + y_{i+1})(x_i y_{i+1} - x_{i+1} y_i) \tag{2}$$

Hence, the point (C_x, C_y) is then detected as the center of OD in the proposed method.

2.3 *Major Vessels by Measuring Width*

Leaving a certain pixel distance from the center of OD detected from the previous stage, a rectangular window boundary is formed surrounding it. Scanning the window boundary clockwise determines the width of the major vessels, shown in Fig. 3b.

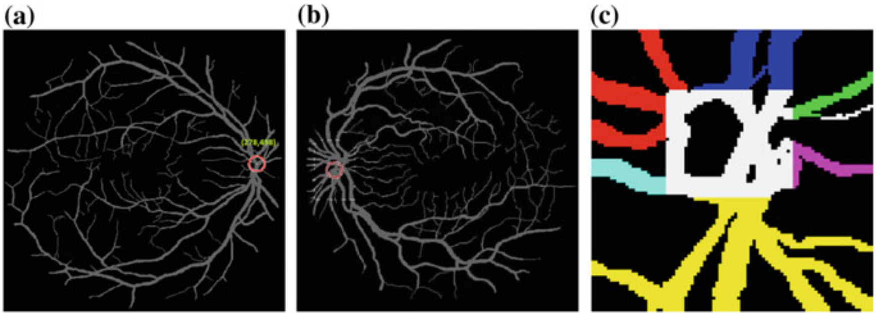


Fig. 3 Sequence of the major vessel width measurement around OD is shown **a** center of OD detected, **b** window boundary around center of OD is drawn, **c** major vessels are identified

Number of white pixels within a black-to-white and white-to-black transition pair shows the width of the major vessels. A structural analysis on the number of widths is done and an assumption on the average width is also observed.

2.4 Vessel Coloring

The same window boundary method described earlier is followed here to color the major vessels. At the outer boundary of the window, clockwise, white pixels are colored until white-to-black transition occurs. The process continues till there is no white pixel left at the outer boundary of the window [7]. Keeping detected center of optic disc at the middle, major vessels with specific width are identified by different colors, see Fig. 3c.

3 Validation Stage

Regular monitoring of the diabetic patients helps to determine the early signs of PDR. Here, two-stage validation starts by passing the fundus image of the patient through the preprocessing and segmentation phases again. Abrupt changes in the count of vessels first, asserts the abnormal presence of newly generated minor vessels. The sudden emanation of tiny vessels is caught by calculating the width of them at the second stage of validation. Finally, the presence of tiny vessels below an empirically derived threshold value is not given colors as they have failed to satisfy the coloring criteria. So, early detection of PDR by two-stage validation method helps in preventing the disease before it aggravates.

The images in Fig. 5c, d show the unusual changes in the region around optic disc center which is indicative of PDR. The proposed method can identify the generated

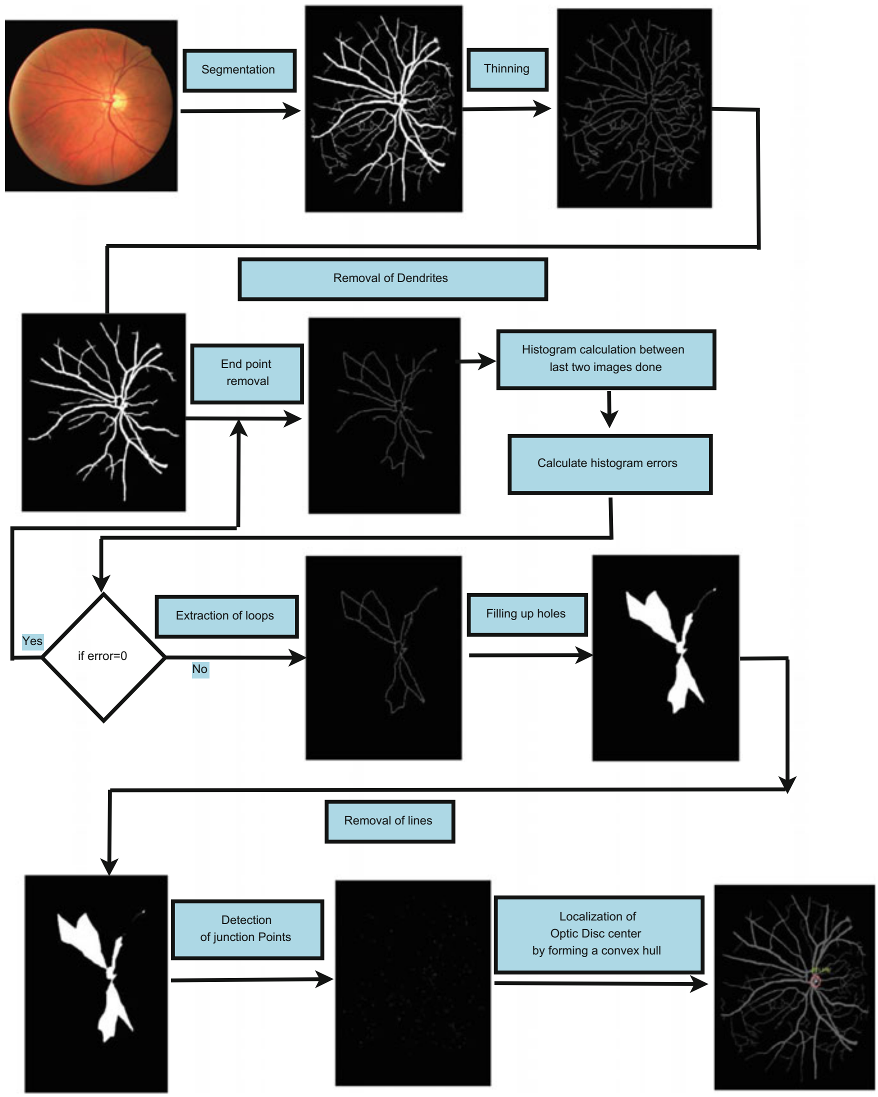


Fig. 4 Flow for optic disc center detection

tiny vessels by measuring their width and not coloring them if they failed the coloring criteria, for not being a part of major vessel. Also, the method triggers an alarming situation by finding sudden mismatch in vessel count around OD. Abrupt changes in OD area proposed in this work thus detects the early signs of PDR.

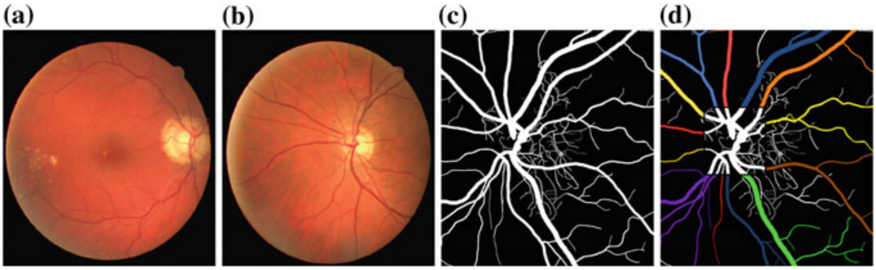


Fig. 5 Images **a** and **b** from DRIVE database are found to be normal here, but images **c** and **d** from STARE database show proliferation of fragile vessels around optic disc

4 Experimental Evaluation

4.1 Materials

Due to low prevalence of PDR images, the proposed method was evaluated on the images collected from DRIVE [5] and STARE [8] databases. The collected RGB images are passed through the several stages of preprocessing to make the images ready for the experiment. Segmentation plays a crucial role here which is done on every image by following the method described in image preprocessing and segmentation section.

4.2 Performance Measures

It is a common practice to measure the performance of an algorithm based on its sensitivity, specificity, and accuracy calculation. Sensitivity, defined in formula 3, measures the proportion of actual positives which are correctly identified. Specificity is the proportion of negatives which are correctly identified and defined in formula 4.

$$\text{Sensitivity} = \frac{\text{TN}}{\text{TN+TP}} \tag{3}$$

$$\text{Specificity} = \frac{\text{TP}}{\text{TP+TN}} \tag{4}$$

Accuracy is calculated as

$$\text{Accuracy} = \frac{(\text{TP+TN})}{(\text{TP+FP+TN+FN})} \tag{5}$$

where TP is true positive, TN is true negative, FP is false positive, and FN is false negative.

Table 1 Results of databases

Database	No. of images	Accuracy	Sensitivity	Specificity
STARE	25	0.88	0.86	0.87

Table 2 Comparison with other PDR detection methods

Methods	Accuracy	Sensitivity	Specificity	Level
Agurto [9]	0.94	0.96	0.83	Image
Goatman [10]	0.91	0.84	0.85	Image
Hassan [3]	0.70	0.63	0.89	Pixel
Jelinek [11]	0.90	0.94	0.82	Image
Welikala [12]	0.97	1	0.90	Image
Proposed method	0.88	0.86	0.87	Image

4.3 Results

We observe that the proposed method accomplishes the task of early signs of PDR detection in the accuracy range between 0.82 and 0.91 and achieves low false positive rates. Table 1 shows the overall result in various databases in average case and Table 2 compares it with many other existing methods. Figure 5a, b shows the experimental results on images from DRIVE [5] dataset. On both the images, the width of major vessels is computed and shown as green lines on the RGB images. At the same time, count on major vessels is also done. Since there is no abnormality found on the images, the images are declared as normal eye images. On the other hand, we found an abnormality on the images shown in Fig. 5c, d in vessels count, the presence of some tiny vessels and moreover, few vessels are not colored as they failed in the coloring criteria (Fig. 5d). We, therefore, declare them as PDR images where proliferation has already started.

5 Conclusion

PDR is the advanced stage of DR where new abnormal blood vessels form in different regions of the retina. In this work, we are focussing on the NVD where fragile vessels grow around the optic disc. DR cannot be cured completely. Photocoagulation (laser analysis) is very effective if it is done before the disease adversely harms retina. Surgical elimination of vitreous gel (vitrectomy) helps in improvement of vision, provided that the massive damage in retina has not been done. An anti-inflammatory medicine or antivascular endothelial growth factor medication injection is helpful in new blood vessel contraction process. This study helps in detecting the early signs of PDR. Also, an initiative has been taken to find the way out to detect the disease at its early stages to prevent from permanent vision loss. This work will be useful for the technical persons and an automated approach of this study will save many diabetic patients from blindness due to diabetic retinopathy. Since no symptom is seen in PDR until the disease turns into the stern, regular observation on diabetic patient's eye is essential as a preventive measure against vision loss.

References

1. Akram MU, Khalid S, Tariq A, Javed MY (2013) Detection of neovascularization in retinal images using multivariate m-mediods based classifier. Comput Med Imaging Graph 37(5):346–357
2. Oloumi F, Rangayyan RM, Ells AL (2012) Computer-aided diagnosis of proliferative diabetic retinopathy. International conference of the IEEE engineering in medicine and biology society (EMBC) 2012:1438–1441
3. Hassan SSA, Bong DB, Premsenthil M (2012) Detection of neovascularization in diabetic retinopathy. J Digit Imaging 25(3):437–444
4. Srivastava R, Wong DW, Duan L, Liu J, Wong TY (2015) Red lesion detection in retinal fundus images using frangi-based filters. In: IEEE engineering in medicine and biology society (EMBC), pp 5663–5666
5. The DRIVE database, Image sciences institute, university medical center Utrecht, The Netherlands. http://www.isi.uu.nl/Research/Databases/DRIVE/. Last accessed on 7th July 2007
6. Gonzalez RC, Eugene Woods R Digital image processing book, 3rd edn. Paperback Publishers
7. Dutta Roy N, Someswar M, Dalmia H, Biswas A (2014) Identification of distinct nerves in retinal fundus images. compImage'14, P.A. Pittsburgh, USA
8. Boyd J (1996) STARE software documentation: diskOptic disk locator. Vis Comput Lab Dept Elect Comput Eng Univ California
9. Agurto C, Honggang Y, Murray V, Pattichis MS, Barriga S, Bauman W et al (2012) Detection of neovascularization in the optic disc using an AMFM representation, granulometry, and vessel segmentation. In: Annual international conference of the IEEE, engineering in medicine and biology society (EMBC), 2012, pp 4946–4949
10. Goatman KA, Fleming AD, Philip S, Williams GJ, Olson JA, Sharp PF (2011) Detection of new vessels on the optic discusing retinal photographs. IEEE Trans Med Imaging 30:972
11. Jelinek HF, Cree MJ, Leandro JJ, Soares JV, Cesar RM Jr, Luckie A (2007) Automated segmentation of retinal blood vessels and identification of proliferative diabetic retinopathy. J Opt Soc Am Opt Image Sci Vis 24:1448–1456
12. Welikala R, Dehmeshki J et al (2014) Automated detection of proliferative diabetic retinopathy using a modified line operator and dual classification. Comput Methods Programs Biomed 114(3):247261

Nilanjana Dutta Roy is a researcher at Indian Institute of Engineering Science and Technology, Shibpur, with specialisation in the field of Image processing and is an Assistant Professor at Institute of Engineering & Management, Kolkata.

Arindam Biswas graduated from Jadavpur University, Kolkata, India, and received his masters and doctorate degree both from the Indian Statistical Institute, Kolkata, India. He is currently Associate Professor in the Department of Information Technology, Indian Institute of Engineering Science and Technology, Shibpur, India. His research interests include digital geometry, image processing, approximate shape matching and analysis, medical image analysis, natural language processing, and biometrics. He has published over 75 research papers in international journals, edited volumes, and refereed conference proceedings and holds one US patent.

Finding Center of Optic Disc from Fundus Images for Image Characterization and Analysis

Nilanjana Dutta Roy and Arindam Biswas

Abstract An automated method for center reference point extraction from retinal fundus images is essentially required for an untroubled image mapping in medical image analysis, image registration, and verification. This paper proposes a spadework, revealing a distinct reference point within optic disc in blood vessel structure of the human eye, analysis on which would serve as an efficient preventive measure for any ocular disease and would strengthen the image verification method along with other extracted features of the human eye at low cost. The proposed method includes segmentation from colored fundus images followed by removal of thin and tiny blood vessels which carry very less information. Removal process comes up with a few bounded polygonal structures, named ring near optic disc. From the named structures near optic disc, a cluster of junction points have been found with maximum members and we made a convex hull out of them. Finally, calculating the centroid of the formed convex hull unveils the center of the optic disc. Experiments are done on some publicly available databases called DRIVE, STARE, and VARIA. Experimental results compared to other standard methods are available in the literature.

1 Introduction

Digital image analysis on retinal images offers huge potential benefits by automated analysis process. In a research setting, an automated method exhibits its ability to examine a large number of images within marginal time and cost and diagnoses deformities more precisely than traditional observation driven techniques. Unnatural behavior in retinal reference features may be associated with the occurrence of

N. D. Roy (✉)
Department of Computer Science and Engineering,
Institute of Engineering and Management, Kolkata, India
e-mail: nilanjanaduttaroy@gmail.com

A. Biswas
Department of Information Technology,
Indian Institute of Engineering Science and Technology, Shibpur, Howrah, India
e-mail: barindam@gmail.com

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_135

retinopathies or cardiovascular diseases. To observe any abnormality in the human retina as a preventive measure, study and analysis on optic disc (OD) are essentially important. On the other hand, due to the overwhelming demand of biometric authentication against vulnerable threats in traditional security systems, strengthening it from every aspect at marginal cost has become an essential work. Optic disc is a key reference for many recognition algorithms [1, 2]. It has been asserted by expert clinicians and scientists that the structure of its blood vessels are distinct in every individual, even for the identical twins and it remains same during his life. The statement which, although literally true, is nevertheless a challenging task to prove technically. OD is the brightest region of any retinal fundus image from where all the blood vessels emanate. It is often considered as a landmark and reference for the other features of any retinal fundus image. An automatic and efficient detection of it plays a significant role for various reasons, like to locate other anatomical components in retinal images, for vessel tracking to diagnose many eye diseases even related to diabetes and for registering the image for personal authentication. Moreover, to find abnormal structures in a retinal image successfully, it is often necessary to mask out the normal anatomy from the analysis. Also, the position of OD can be used as a reference length for measuring distances in retinal images, especially from the location of macula. Localization of OD is hence an essential and challenging work. OD extraction is quite difficult since its brightness, color, and contrast of OD is similar to other components such as cotton wools and exudates [3]. Most of the algorithms to localize OD fail when pathological regions exist in retina images [4, 5]. Some other algorithms are quite expensive and suffer from high computational cost [6–8]. We have taken an initiative in support to the undeniable truth by extracting a key feature from the vascular structure, along with other invariable features present in it which would be further helpful in image registration and verification process. Hence, it is a well-suited identification characteristic for biometrics also. Therefore, identifying the right parameter plays a significant role in the process to provide a stringent and unambiguous authentication in marginal time. So keeping these situations in mind, we are trying to explore a unique reference key, say, center reference key of OD which would be robust enough to work in any situation at normal cost and thus will help in biometric template matching in future.

2 Proposed Method

The proposed method is fundamentally based on two major phases. Phase I narrates the preprocessing phase which includes segmentation from the colored retinal images. Phase II describes a way to locate the center reference point of OD. Removal of thin and tiny blood vessels called dendrites initiates the Phase II process followed by finding the centroid of a cluster points with maximum neighboring points. The below-mentioned steps further help us to define the proposal more precisely.

2.1 Vessel Segmentation and Image Preprocessing

The process starts with a segmentation of the images using a previously developed approach that uses supervised pixel classification with a Gaussian filter set and classification by a k-nearest neighbor classifier [9]. The resulting image represents the likelihood of each pixel belonging to a vessel. In order to trace the vessel path and obtain structural mapping, a connected binary vessel image is required which may be obtained using a vessel reconnection algorithm based on a graph search [10]. The binary vessel image is generated from the vessel probability image using Otsus thresholding method [11]. The Otsu threshold minimizes the intra-class variance for the foreground and the background classes which is basically vessels and non-vessel region. Then, the skeleton of blood vessels is obtained by applying mathematical morphology which reduces the vessel to a center line of single pixel width [12].

While scanning the segmented image from its top-left to bottom-right corner, the width [13] of each blood vessel have been calculated on each point upon a threshold value 5 pixels. If the width is found to be lesser than the threshold, that vessel is eliminated from the original image considering it as a dendrite.

2.2 Localization of OD Center

Keeping the importance of localizing OD in mind, we have taken an initiative to locate an exact point, the center of it within the white bright region called OD. This robust method will successfully work with low-resolution images, in presence of pathological regions and even in infected images. As the algorithm proceeds, the center of OD region has been found following the below-mentioned method, see Fig. 1 to have a concise view of the process. Figure 2b exhibits the ultimate outcome of this section from a segmented image. It has been statistically found from the binary images, that there is a normal tendency of the major blood vessels near optic disc to bifurcate and cross each other which further form a few polygonal loop-like structures, called rings. We are emphasizing on the rings here by considering them as a trademark to locate the optic disc in this algorithm. But overlapping on dendrites may also appear as same which carries redundant information. So, removing dendrites is an essential task to keep the loops near OD intact.

A morphological thinning operation is applied on the segmented fundus image to get a thinned, skeleton view of it initially, shown in Fig. 3b. Next, we tried to reveal all the rings present in segmented fundus image. Then, to remove the tiny vessels, a scan at each point of the segmented input image is being done to measure the width of all existing vessels. For any vessel, having its width as less than a threshold value of 5 pixels, is considered as dendrite and the same is removed from the original segmented image, see Fig. 3c. Here, we tried to remove all the end points or tails of blood vessels which would end up at any bounded region or polygonal loops. A white pixel whose seven neighbors are black in a 3×3 scanning window has

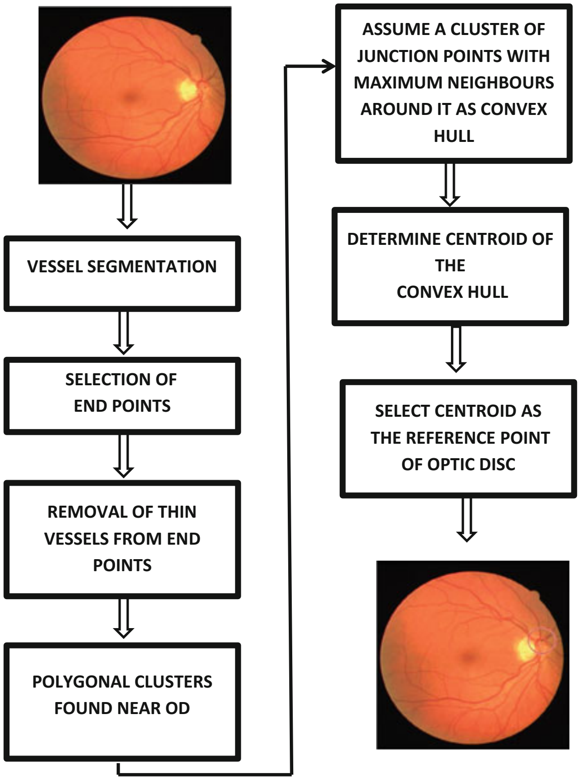


Fig. 1 Overview of the method

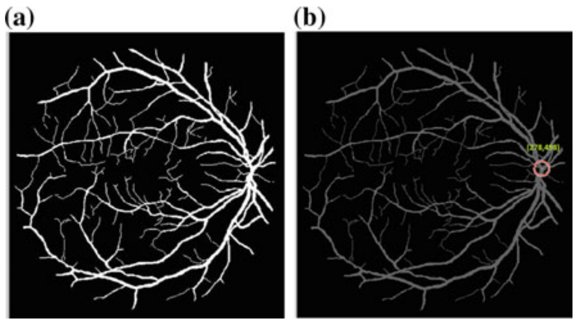


Fig. 2 **a** Segmented fundus image. **b** Optic disc located

been assumed as end point in this method. This comes within an iteration of few steps. At each step, end points are removed from the thinned image by making them same as background and it is being compared with the histogram of thinned image obtained in the previous iteration. Comparison is made following the histogram error calculation between both the images at every iteration. Error zero indicates that both the images are exactly same by anatomy, whereas some error shows that there are further scopes exist yet for removal of its tails. This process stops when no end points

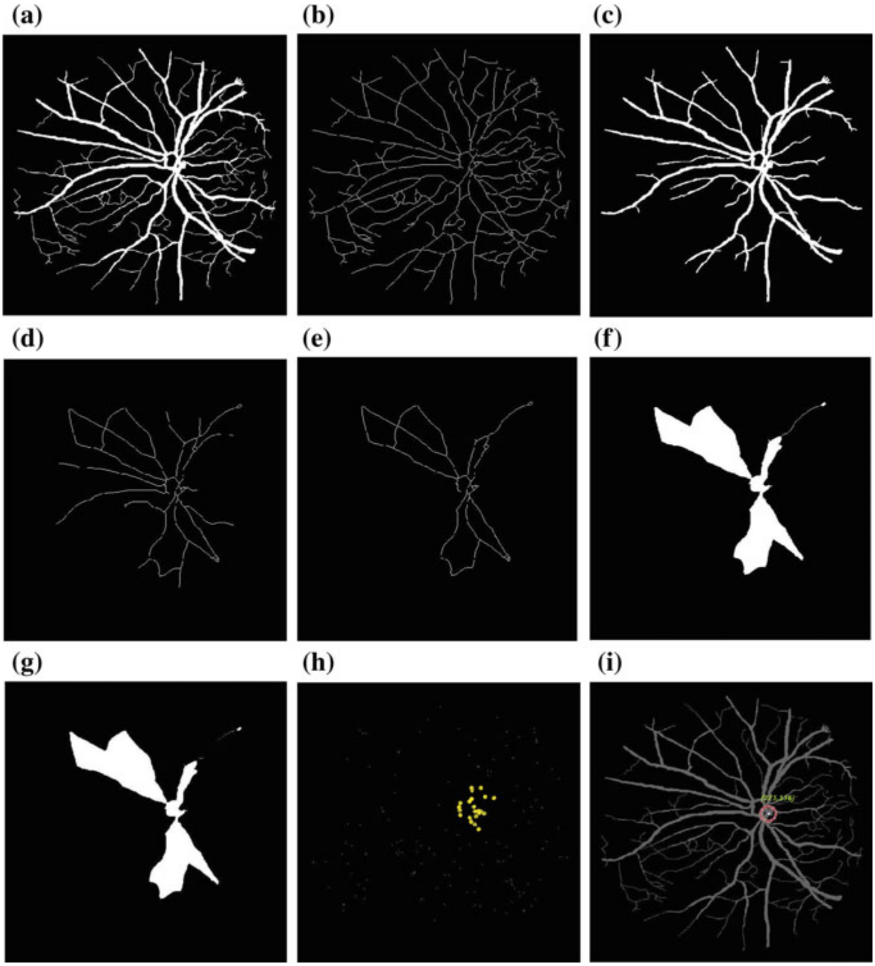


Fig. 3 **a** Segmented fundus image. **b** Thinned image obtained after applying morphological function. **c** Image after removing the dendrites. **d** Image after a few iteration. **e** After extracting the loops in the image. **f** After applying area fill function to bridge the holes. **g** Further removal of lines connecting the loops. **h** All junction points detected. **i** Optic disc located

can be further removed and the loops have been found. Following this way, all the end points are removed from the thinned images by making every eight neighborhood pixels as black, refer Fig. 3e. Histogram error 0 says both the images are found to be identical. On the flip side, the histogram error not equal to zero indicating not identical images, signifies that the end points could further be removed from the images and the process continues till the next match comes. Single line removal process stops at every junction and checks for its neighborhood regions.

2.3 Determining Convex Hull Out of Junction Points Near OD

After all the single line pixels have been removed, it turns into a smallest possible structure with all the polygonal rings, see Fig. 3e. We have used an algorithm for region filling which performs on background pixels of the binary image, starting from any seed pixel within the bounded region and covers all the regions within the bounded loops. After the polygonal holes have been filled up [14], see Fig. 3f, a scanning window of size 3×3 moves around horizontal direction to determine a polygon with maximum number of single lines or neighbors connected with it. Neighbor selection is again done in anticlockwise direction and by connected component analysis around the bounded region. All single lines attached with the region are now removed by making them same as background pixel. Finally, we are now left with a small colored region and the junction points [15] within that bounded region has been found. We are concerned about those selected junction points in this region as they would help us to accomplish the goal of determining the center of OD. We now try to convert those scattered points into a convex hull or minimal convex polygonal (MCP) shape to locate the centroid of the bounded area.

Converting a few points into a convex hull starts by finding the bottom most point, say P in this case. Remaining points are sorted in ascending order with respect to their angles formed with the bottom most point P. Traversing counterclockwise with left turn for the next node is safe, whereas taking right turn may discard the next point to be chosen from the remaining list saying the initial and final points are out of boundary line. The process ends after meeting the initial point from where it has been started. The ultimate junction points near OD that we have generated from the earlier steps are now being converted into a convex hull or MCP by following the Graham Scan method [16].

2.4 Centroid of MCP as Center of OD

Once the focused region has been found, it becomes an effortless job to calculate the centroid of the MCP or the convex hull. Let us assume the bounded region is a polygon with N vertices and $N-1$ line segments; (x_i, y_i), where $i = 0$ to $N-1$. The last vertex (x_N, y_N) is assumed to be the same as initial vertex (x_i, y_i), marking it as a closed polygon. The area of this region is then decided by

$$A = \frac{1}{2}\sum_{i=1}^{N-1}(x_i y_{i+1} - x_{i+1} y_i) \tag{1}$$

Area further helps to calculate the centroid of the bounded region where (x_N, y_N) is assumed to be same as (x_0, y_0).

$$C_x = \frac{1}{6A} \sum_{i=1}^{N-1} (x_i + x_{i+1})(x_i y_{i+1} - x_{i+1} y_i) \quad (2)$$

and

$$C_y = \frac{1}{6A} \sum_{i=1}^{N-1} (y_i + y_{i+1})(x_i y_{i+1} - x_{i+1} y_i) \quad (3)$$

Hence, the point (C_x,C_y) is then considered as the center of OD in our proposed method.

3 Results and Discussion

The purpose of this spadework was to identify the center pixel of OD in the human retina. Finding center of OD plays a significant role in many aspects as discussed earlier. Additionally, it could act as a root of a retinal vascular tree generated from the retinal vascular network. Any geometric transformation would successfully be applied here for future applications. This major extracted feature could again be used as an important biometric parameters for preparing retinal template for image registration by matching digital images. Besides, any deformities, like retinal dislocation etc., could easily be detected for medical image analysis. This section presents ultimate outcomes of both Phase I and Phase II within a single block (Table 1).

In literature, there are contributions from many other authors also in the similar area. We tried to reveal a comparative study which proves our proposed method is comparable with some of the other existing works, please see Table 2. Figure 4 shows the detected OD center on different images of DRIVE [17], VARIA [18] and

Table 1 Experimental results showing location of OD center on DRIVE database

DRIVE database							
Image	Optic disc	Image	Optic disc	Image	Optic disc	Image	Optic disc
1	(248, 84)	11	(254, 63)	21	(254, 69)	31	(263, 377)
2	(248, 484)	12	(255, 76)	22	(283, 475)	32	(266, 515)
3	(298, 71)	13	(275, 497)	23	(251, 417)	33	(257, 512)
4	(282, 346)	14	(283, 487)	24	(287, 481)	34	(281, 468)
5	(279, 78)	15	(249, 229)	25	(298, 466)	35	(267, 77)
6	(255, 470)	16	(272, 491)	26	(248, 86)	36	(279, 482)
7	(252, 483)	17	(280, 470)	27	(283, 500)	37	(274, 496)
8	(282, 491)	18	(249, 501)	28	(269, 486)	38	(290, 499)
9	(268, 89)	19	(276, 490)	29	(289, 505)	39	(282, 89)
10	(281, 466)	20	(264, 501)	30	(263, 469)	40	(282, 486)

Table 2 Comparison between different methods for OD localization

Results of other methods to detect OD					
Algorithms	Result (%)	Data set	Running time (s)	Distance (pixels)	System configuration
Youssif	100	DRIVE	210	17	Intel Core 2 Duo 1.7 GHz and 512 Mb RAM
Sekhar	85	DRIVE	NA	NA	NA
Rangayyan	100	DRIVE	2294	23.2	Intel Core 2 Duo 2.5 GHz and 1.96 GB RAM
Zhu et al.	90	DRIVE	NA	18	NA
Park et al.	90.25	DRIVE	4	NA	Intel 1 GHz and 1 GB RAM
Dehghani	100	DRIVE	27.6	15.9	Intel Core 2 Duo 2.67 GHz and 3.24 GB RAM
Proposed method	92	DRIVE	20	15	Intel Core 2 Duo 2.67 GHz and 2 GB RAM

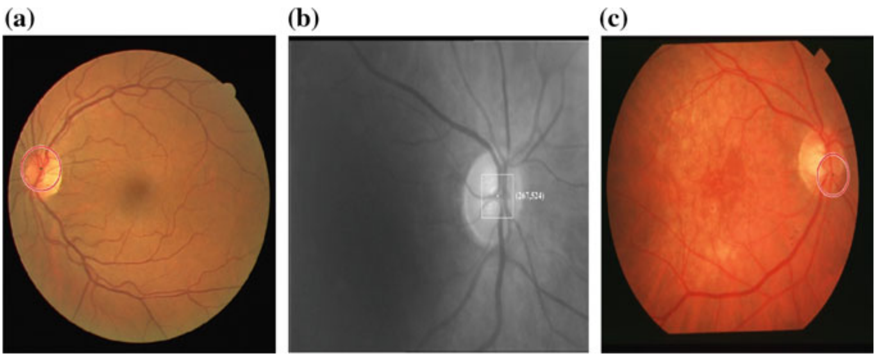


Fig. 4 Results of the proposed method on various databases. **a** Detected center of OD on DRIVE image. **b** Detected center of OD on VARIA image. **c** Detected center of OD on STARE image

STARE databases. The images of DRIVE have been captured using Canon CR5 non-mydriatic 3 CCD cameras with a 45° FOV for medical imaging.[1]

4 Conclusion

Characterization of several features from the human retina is immensely appreciable work. Diagnosis of any ocular diseases need a thorough analysis on retinal fundus images, especially on OD and bifurcation points. Also, with the rapid growth of biometric authentication, searching an efficient feature is essentially important and

[1]The results may vary a bit due to different clarity of the images published by other databases.

challenging task for faster computation. So keeping these primary objectives in mind, i.e., analysis of specific features for disease detection, for characterization, and to search new features for preparing a retinal template for biometric authentication, we have executed the proposed method on 32 bit MATLAB 2013a version on 32 bit Windows XP running on Intel Core-2 dual processor with 2 GB of RAM. Locating center of OD is essentially important for finding the accurate distance between Macula and OD. The experimental results are clearly showing the detected OD locations on segmented fundus images from various databases.

References

1. Farzin H, Abrishami Moghaddam H, Moin MS (2008) A novel retinal identification system. EURASIP J Adv Sig Process 2008. Article ID 280635. https://doi.org/10.1155/2008/280635
2. Ortega M, Penedo MG, Rouco J, Barreira N (2009) Retinal verification using a feature-point based biometric pattern. EURASIP J Adv Sig Process, Article ID 235746. https://doi.org/10.1155/2009/235746
3. Sumathy B, Poornachandra S (2013) Feature extraction in retinal fundus images. In: IEEE Xplore proceedings of information communication and embedded systems (ICICES), Chennai
4. Osareh A (2004) Automated identification of diabetic retinal exudates and the optic disc. Ph.D. dissertation. Department of Computer Science, Faculty of Engineering, University of Bristol, Bristol, UK
5. Li H, Chutatape O (2001) Automatic location of optic disc in retinal images. Proc Int Conf Image Process (ICIP) 2:837–840
6. Youssif AA, Ghalwash AZ, Ghoneim AS (2008) Optic disc detection from normalized digital fundus images by means of a vessels direction matched filter. IEEE Trans Med Imag 27:11–18
7. Rangayyan RM, Zhu X, Ayres FJ, Ells AL (2010) Detection of the optic nerve head in fundus images of the retina with Gabor filters and phase portrait analysis. J Digit Imag 23(4):438–453. https://doi.org/10.1007/s10278-009-9261-1
8. Zhu X, Rangayyan RM, Ells AL (2010) Detection of the optic nerve head in fundus images of the retina using the hough transform for circles. J Digit Imag 23(3):332–341. https://doi.org/10.1007/s10278-009-9189-5
9. Niemeijer M, Staal J, Ginneken B, Loog M, Abramoff M (2004) Comparative study of retinal vessel segmentation methods on a new publicly available database. Proc SPIE 5370:648–656
10. Joshi V, Garvin M, Reinhardt J, Abramoff M, Identification and reconnection of interrupted vessels in retinal vessel segmentation. In: IEEE, ISBI, image segmentation methods, vol FR-PS3a.7, pp 1416–1420
11. Otsu N (1979) A threshold selection method from gray-level histograms. IEEE Trans Sys Man Cyber 9(1):62–66
12. Rockett P (2005) An improved rotation-invariant thinning algorithm. IEEE Trans Pattern Anal Mach Intell 27:1671–1674
13. Lowell J, Hunter A, Steel D, Basu A, Ryder R, RL Kennedy (2004) Measurement of retinal vessel widths from fundus images based on 2-D modeling. IEEE Trans Med Imaging 23(10)
14. Hearn D, Pauline Baker M Computer graphics, 3rd edn. Paperback Publishers
15. Saha S, Dutta Roy N (2013) Automatic detection of bifurcation points in retinal fundus images. Int J Latest Res Sci Technol 2(2):105–108. ISSN (Online):2278-5299. http://www.mnkjournals.com/ijlrst.htm
16. Graham RL (1972) An efficient algorithm for determining the convex hull of a finite planar set. Inf Process Lett 1:132–133
17. The DRIVE database, Image sciences institute, university medical center utrecht. The Netherlands. http://www.isi.uu.nl/Research/Databases/DRIVE/. Last accessed on 7th July 2007
18. VARIA Database, Department of Computer Science of the Faculty of Informatics of the University of Corua, http://www.varpa.es/varia.html

Nilanjana Dutta Roy is a researcher at Indian Institute of Engineering Science and Technology, Shibpur, with specialisation in the field of Image processing and is an Assistant Professor at Institute of Engineering & Management, Kolkata.

Arindam Biswas graduated from Jadavpur University, Kolkata, India, and received his masters and doctorate degree both from the Indian Statistical Institute, Kolkata, India. He is currently Associate Professor in the Department of Information Technology, Indian Institute of Engineering Science and Technology, Shibpur, India. His research interests include digital geometry, image processing, approximate shape matching and analysis, medical image analysis, natural language processing, and biometrics. He has published over 75 research papers in international journals, edited volumes, and refereed conference proceedings and holds one US patent.

A Robust Method for Image Copy-Move Passive Forgery Detection with Enhanced Speed

Asif Hassan and V. K. Sharma

Abstract Forgery detection of images is presently one of the fascinated research fields. Copy-move forgery is the most commonly used methods for image forgery. A novel method is proposed in this paper, which is an effective and advanced method for detecting copy-move forgery. The proposed method is a block matching technique with reduced computational speed and less computational complexities. The efficiency of outcome is also improved. The image is segmented into fixed dimensions of overlying blocks and then discrete cosine transform (DCT) is applied to each block to extract its features. Then, the mean of each block is obtained. The mean of each block is compared with other blocks to find the similarity between the blocks. The computational outcomes are shown that indicates the proposed method is robust to detect copy-move forgery efficiently with enhanced speed.

1 Introduction

An image is an artifact that represents visual insight of an event. We live in a digital world where digital images are used as a means of communication. Images are everywhere, on the Internet, social media, newspapers, etc. The authenticity of images is in question because it is very easy to edit any image using easily available image manipulating tools [1–4]. Hence, digital image forgery detection is important to authenticate the images. Image authentication methods can be classified into two modules such as active methods and passive methods. Active method is the technique in which prior information about the original image such as watermarking or signature [5–7] which is embedded inside the image is known for forgery detection. It is a drawback because in various situations prior information about the image is not available [8–11].

A. Hassan (✉) · V. K. Sharma
Bhagwant University, Ajmer, Rajasthan, India
e-mail: asif.43hassan@gmail.com

V. K. Sharma
e-mail: viren_krec@yahoo.com

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_136

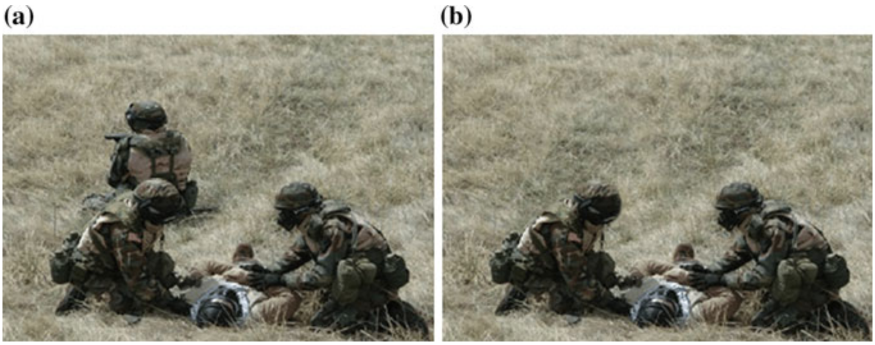


Fig. 1 Copy-move forgery: **a** Original image, **b** Copy-move forged image

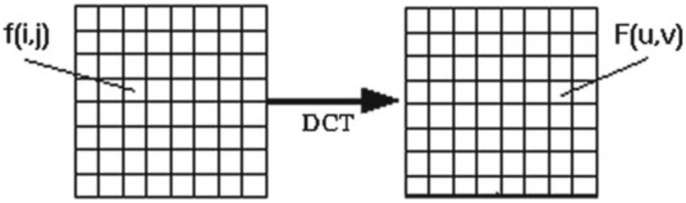


Fig. 2 Discrete cosine transforms

The Passive method is also known as the *blind technique* in which information regarding the image is not available [12, 13]. Hence, this method authenticates the image without the existence of the original information.

One of the most commonly used methods of image altering is hiding a region in the image and distorting the image information. Copy-move forgery is the most common image forgery used to hide the portion of the image. In copy-move forgery, some portions of an image are copied and posted on another area of the same image. An example of copy-move forgery is shown in Fig. 1.

The most frequently used method for image copy-move detection is block matching in which the image is divided into the same size overlying blocks then feature of each block is extracted and then each block is compared with other blocks in the same image. At last, the outcome of forgery detection is decided based on matched block features. During the feature extraction process, the essential features are chosen from the blocks using a discrete cosine transform (DCT)[1, 3]. These essential features are used to compare the blocks. DCT helps to segment the image or spectral sub-bands of differing importance corresponding to the image visual quality. The DCT is like the discrete Fourier transform which transforms the image into the frequency domain from the spatial domain as shown in Fig. 2.

Fadl and Semary (accelerated method) [1] proposed a method using K-means classifier and images of 128×128-pixel grayscale images. The proposed method works on direct computation, without using the classifier, for an image size of any pixel with reduced computational steps.

The paper is organized as follows: Sect. 2 presents the proposed system in details. Section 3 presents the test and results and Sect. 4 is the conclusion.

2 Proposed Method

The core of the proposed method is to examine whether the input image contains copied regions or not and to identify the region from where it is copied from the same image.

The proposed method is explained in the following steps:

Step 1 The color image or gray image of any pixel value is taken as input for which copy-move detection must be performed.

Step 2 The Color image is converted from RGB to gray. Gray image is retained as it is.

Step 3 Images are segmented into equal size overlaying blocks. The total number of blocks depends on the pixel of the image. It is calculated using the following equations:

$$\mathbf{Block\ size} = \mathbf{2t} \tag{1}$$

where $\mathbf{t} = \mathbf{log2}\ (\mathbf{M} \times \mathbf{N}) - \mathbf{12}$, "M" is the number of rows, and "N" is the number of columns.

Therefore,

$$\mathbf{the\ total\ number\ of\ blocks} = ((\mathbf{M} * \mathbf{N}))/(\mathbf{Block\ size}) \tag{2}$$

The minimum value of $t = 2$

Step 4 Calculate the DCT for each block.

The general for DCT is

$$F(u, v) = \left(\frac{2}{N}\right)^{\frac{1}{2}} \left(\frac{2}{M}\right)^{\frac{1}{2}} \sum_{i=0}^{N-1} \sum_{j=0}^{M-1} \Lambda(i) \cdot \Lambda(j) \cdot \cos\left[\frac{\pi.u}{2.N}(2i+1)\right] \cos\left[\frac{\pi.v}{2.M}(2j+1)\right] \cdot f(i, j) \tag{3}$$

For the input image N by M, $f(i, j)$ is the intensity of the pixel in row i and column j; $F(u, v)$ is the DCT coefficient in row $k1$ and column $k2$ of the DCT matrix.

Step 5 Calculate the mean of obtained DCT of each block.

Step 6 Then, the mean of each block is matched with all other blocks in the same image.

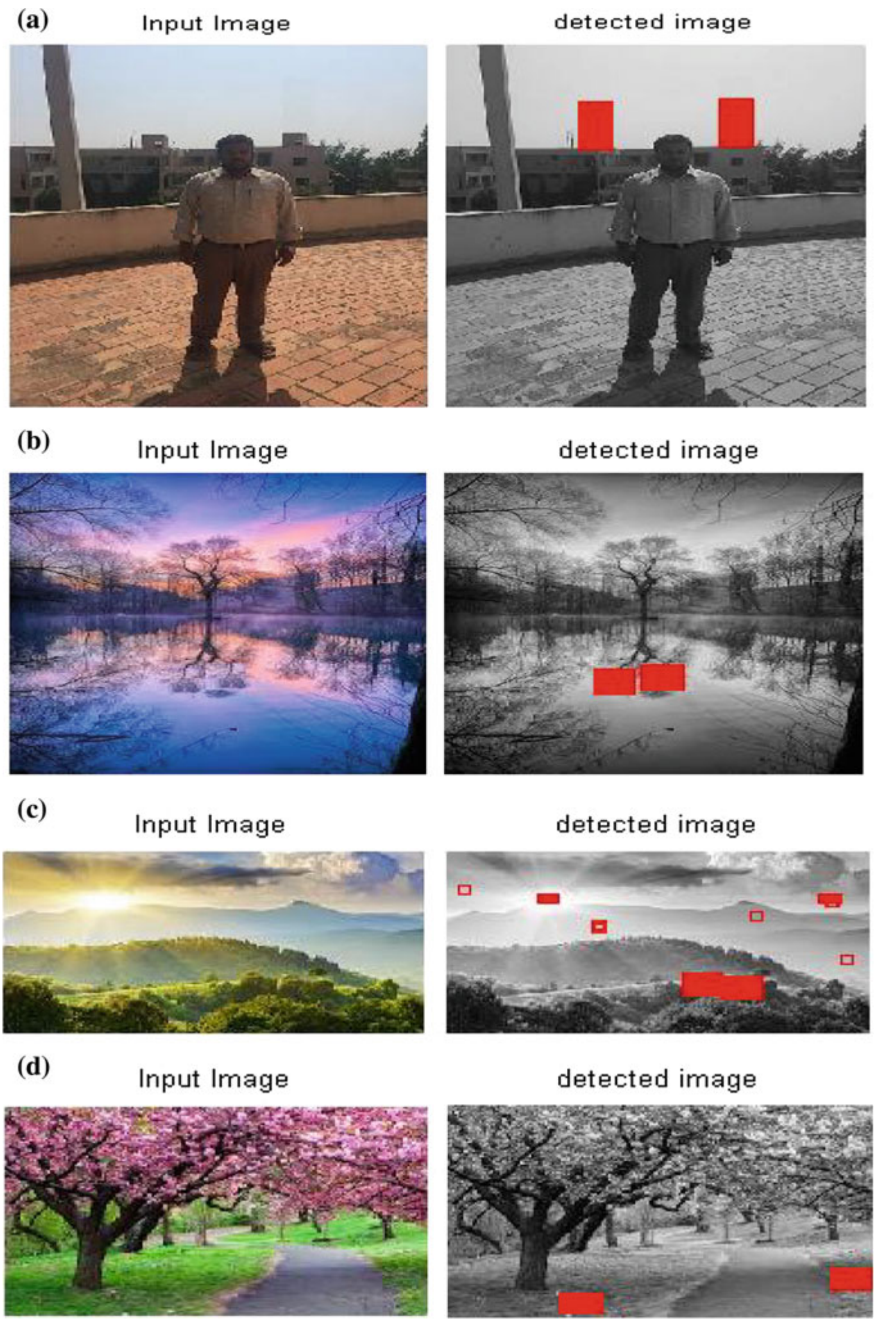


Fig. 3 **a**–**g** Random rectangular region detected

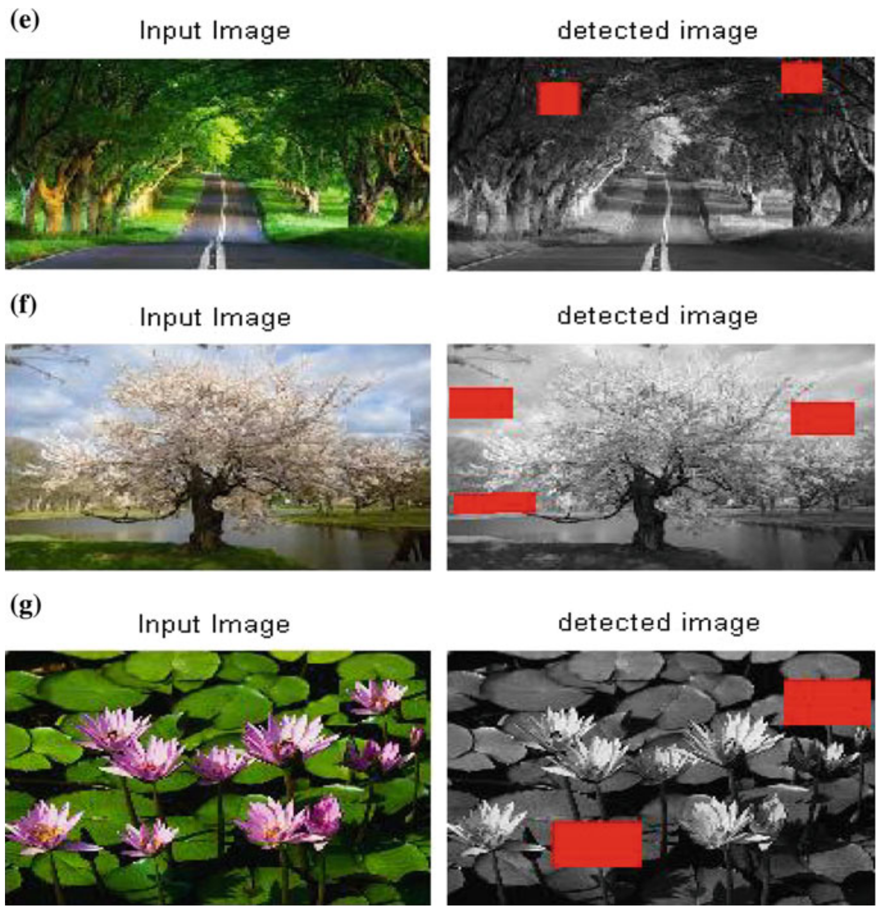


Fig. 3 (continued)

Step 7 If the similarity is found then the block is highlighted (considered as a copied block).

3 Simulation and Results

The experimental results are discussed in this section. The tests were carried out on the MATLAB R2013a, RAM 2 GB, and processor 2.90 GHz, the images with the different pixel values, saved in BMP format, are tested to check the computational speed and robustness of the proposed algorithm.

Figure 3 shows two images: The input image and detected the image. The random rectangular region is copied and pasted onto the same image. The highlighted region represents the copied region and pasted region. The test conducted without having prior knowledge of the images with different pixel values.

Table 1 shows the computational time of forgery detection of random rectangular regions applied on images using the proposed method and method proposed by Fadl and Semary [1] (the accelerated method). As per the table, the proposed method is improved with enhanced speed compared to the accelerated method.

Figure 4 shows that the performance of the accelerated method, which produces noise while detecting higher pixel value images, and hence its performance is reduced compared to the proposed method. As the pixel values are more, the accelerated method produces noise. The proposed method is more accurate even for higher pixel value. This is highlighted in the first row and last row of Table 1.

Table 1 Comparison of computational time between the proposed method and the accelerated method

Figure	Image pixels	Proposed method (s)	Accelerated method (s)
3a	304 × 408	44.99	78.70 (with noise)
3b	183 × 275	13.07	23.76
3c	300 × 168	14.09	21.09
3d	284 × 177	13.09	22.09
3e	284 × 177	13.95	22.60
3f	276 × 183	14.17	22.03
3g	400 × 300	42.23	84.96 (with noise)

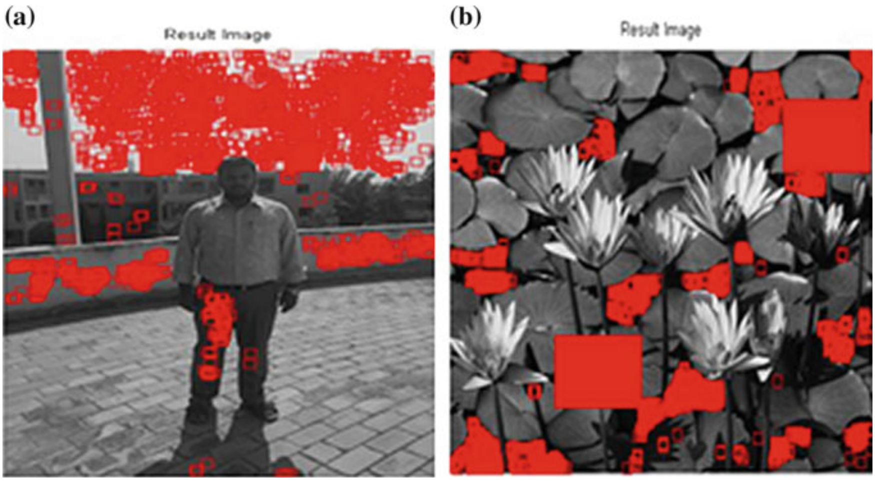


Fig. 4 **a**, **b** Noise during detection in the accelerated method

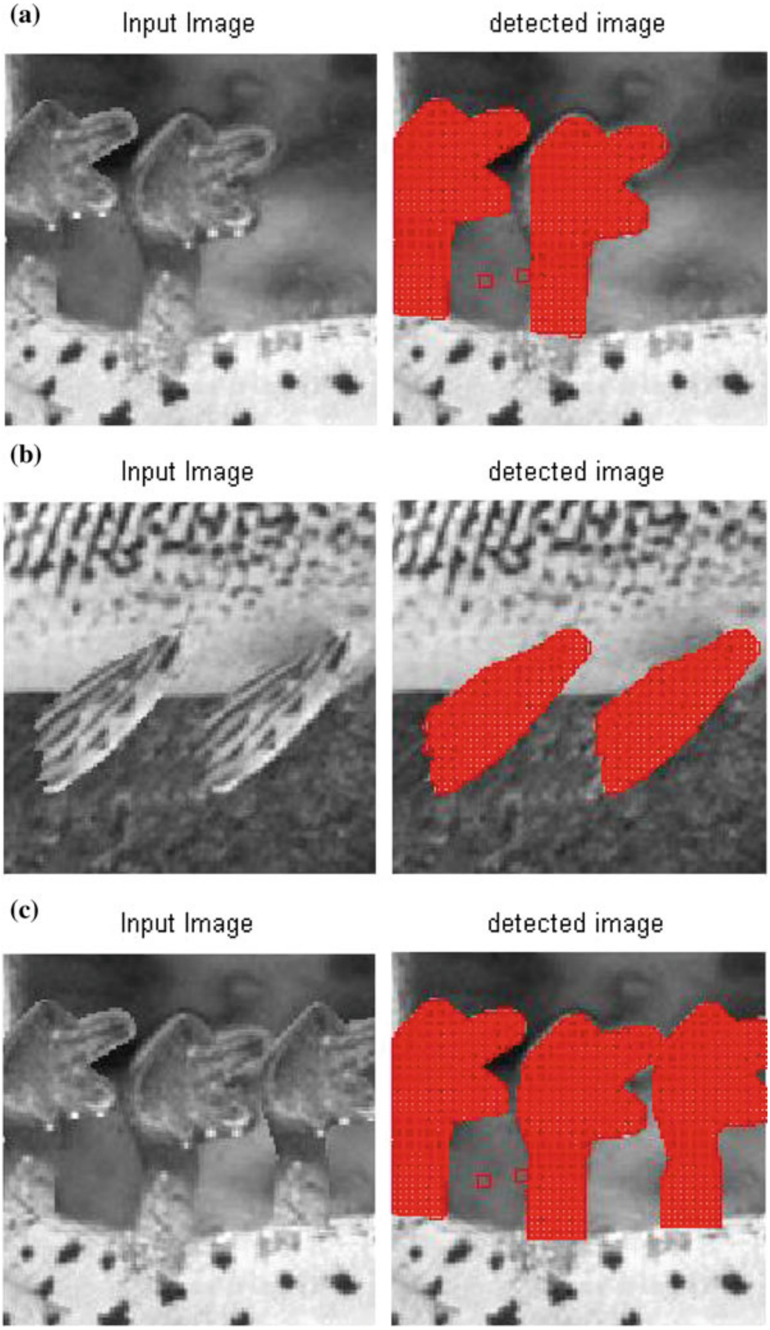


Fig. 5 **a**–**c** Copy-move detection for irregular regions

Table 2 Comparison of computational time between the proposed method and the accelerated method

Figure	Image pixels	Proposed method (s)	Accelerated method (s)
5a	128 × 128	3.17	9.46
5b	128 × 128	3.00	6.42
5c	128 × 128	3.34	6.86

Figure 5 shows an irregular region that is copied and pasted onto the image. The test shows the detection for the single region and multi-regions. All the images are 128 × 128 pixel values.

Table 2 shows the computational time of forgery detection of random rectangular regions applied on images using the proposed method and method proposed by Fadl and Semary [1] (the accelerated method). As per the table, the proposed method is improved with enhanced speed compared to the accelerated method.

4 Conclusion

This work is robust to detect copy-move forgery efficiently, by means of DCT features and taking the mean of DCT of blocks and comparing with other blocks of the same image. It works without any prior information about the image. The proposed work is fast and more effective for any pixel values compared to existing methods.

References

1. Fadl SM, Semary NA (2014) A proposed accelerated image copy-move forgery detection. In: Visual communications and image processing conference, IEEE
2. Lynch G, Shih FY, Liao HYM (2013) An efficient expanding block algorithm for image copy-move forgery detection. Inf Sci 239:253–265
3. Huang Y, Lu W, Sun W, Long D (2011) Improved DCT-based detection of copy-move forgery in images. Forensic Sci Int 206(1):178–184
4. Lin HJ, Wang CW, Kao YT (2009) Fast copy-move forgery detection. WSEAS Trans Sig Process (World Sci Eng Acad Soc) 5(5):188–197
5. Bhargava N, Sharma MM, Garhwal AS (2012) An improved image authentication technique using random-sequence based secret-sharing scheme. In: 2012 International conference on radar, communication and computing (ICRCC), 21–22 Dec. 2012. IEEE 07 Feb 2013
6. Katzenbeisser S, Petitcols FAP (2000) Information techniques for stenography and digital watermarking. Norwood, A, Artec House
7. Alam S, Jamil A, Saldhi A (2015) Digital image authentication and encryption using digital signature. In: 2015 international conference on advances in computer engineering and applications (ICACEA), 19–20 Mar 2015. IEEE 23 July 2015
8. Singh VK, Tripathi RC (2011) Fast and efficient region duplication detection in digital images using sub-blocking method. Int J Adv Sci Technol 35:93–102

9. Liu G, Wang J, Lian S, Wang Z (2010) A passive image authentication scheme for detecting region duplication forgery with rotation. J Netw Comput Appl 34(5):1557–1565
10. Sebe N, Liu Y, Zhuang Y, Huang T, Chang S-F (2007) Blind passive media forensics: motivation and opportunity. In: Multimedia content analysis and mining. Springer, Berlin, pp 57–59
11. Chen C-H, Tang Y-L, Hsieh W-S (2014) Color image authentication and recovery via adaptive encoding. In: 2014 international symposium on computer, consumer and control (IS3C), 10–12 June 2014. IEEE 30 June 2014
12. Zhang Z, Ren Y, Ping XJ, He ZY, Zhang SZ (2008) A survey on passive-blind image forgery by doctor method detection. In: Proceedings of seventh international conference on machine learning and cybernetics, pp 3463–3467
13. Kou G, Ma Y (2015) Color image authentication method based on triple-channel spiking cortical model. In: 2015 10th international conference on broadband and wireless computing, communication and applications (BWCCA), 4–6 Nov 201. IEEE 03 Mar 2016
14. Cox IJ, Miller ML, Bloom JA (2002) Digital watermarking San Francisco. Morgan Kaufmann, Burlington

Feature Extraction and Classification of Epileptic EEG Signals Using Wavelet Transforms and Artificial Neural Networks

Upasana Chakraborty and R. Mary Lourde

Abstract Marked by unpredictable seizures, epilepsy is the fourth most prevailing chronic neural disorder. This neurodegenerative disorder can attack individuals belonging to any category or age group. Also, the resulting seizures can be of any type. There is always a possibility of misjudging the symptoms with psychogenic nonepileptic events. Thus, in addition to the common methods like functional magnetic resonance imaging (fMRI) and positron emission tomography (PET), electroencephalography (EEG) is a useful tool to differentiate epilepsy from other neurodegenerative disorders. However, EEG measures brain activity directly unlike the other two techniques that measure changes in blood flow to a certain part of the brain. Hence, EEG is most widely used. The paper focuses on conversion of time domain brain signals into time–frequency domain using wavelet transforms followed by extraction of various statistical and nonlinear features. These features are then fed to the neurons of an artificial neural network (ANN) which indicates the presence of epilepsy in an individual.

1 Introduction

Epilepsy is the fourth most common neurodegenerative disorder as per the Epilepsy Foundation. It can be characterised by frequent seizures which are unpredictable. Statistically, in the US itself, 1.3–2.8 million people are still fighting with this disorder with small kids and elderly people being the most prone [1]. Since not many people respond to the medication as expected, there is a need to identify the root causes through further research. It can help to identify why such type of seizures arise and if it is possible to prevent or reduce its impact if realised before its onset. Though

U. Chakraborty (✉) · R. Mary Lourde
Department of Electrical and Electronics Engineering, Birla Institute of Technology and Science, Pilani, Dubai Campus, Dubai, United Arab Emirates
e-mail: upasana.chakraborty96@gmail.com

R. Mary Lourde
e-mail: marylr@dubai.bits-pilani.ac.in

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_137

there are several other ways to record the brain activity like fMRI and PET, EEG is the only one that is noninvasive, and hence do not have any harmful effects on the subject. Major part of the EEG signal is concentrated in the frequency range of 0–80 Hz. This range is subdivided into five frequency bands—delta, δ (0–3.5 Hz); theta, θ (4–7.5 Hz); alpha, α (8–13 Hz) and beta, β (14–30 Hz). Above 30 Hz lies the gamma band, γ. The amplitude of brain signals varies from around 10 to 100 μV. EEG signal amplitude of an epileptic person is greater compared to that of a normal person [2]. But, since brain activity changes with age, it can be very difficult to conclude if a person has epilepsy by looking at the amplitude of brain signals in time domain, thus making feature extraction necessary.

In this work, authors used discrete wavelet transforms (DWT) to extract the wavelet coefficients that can help to analyse EEG data in time–frequency domain. This approach unravels various features that remain hidden in only time or frequency domain analysis. Use of wavelet coefficients also helps in reducing the number of terms to deal with the huge time domain data that can be represented with few wavelet coefficients. These coefficients are further used to extract various features. Some are statistical parameters like mean, standard deviation and others are nonlinear features like fractal dimensions [3]. These features are finally fed to the input neurons in an ANN-a tool used for classification. Using a simple algorithm, it can classify whether an individual is suffering from epilepsy or not. Though several attempts have been made to do the same, there is always scope for improvement. All the simulations are performed in MATLAB.

2 Literature Review

In [3], support vector machines (SVM) is used as a classification tool. SVM is trained using certain statistical and nonlinear features. Different SVM classifiers are used for different sub-bands and based on the maximum score, final result is obtained. Using various kernel functions, seizure and seizure-free EEG were classified with a very good accuracy. Only drawback is that SVM fitting and prediction takes longer time compared to ANN. This makes the system complex. Also, kernelised SVMs are nonparametric. Support vectors are selected from among the training set with some bias value so the number of outputs is less compared to the number of inputs but at times, the number of support vectors is equal to the number of training datasets. In such situations, using ANN is a better option as SVM works best when there is only one output.

Guo et al. [4] used relative wavelet energy of frequency bands as elements for classification. EEG signal was divided into several frequency bands out of which relative wavelet energy of two bands—low and middle frequency bands were extracted. This was fed to a three-layered neural network which was used for classification with an accuracy of nearly 95%.

3 Methodology

3.1 Dataset Used

The dataset used was taken from the Centre of Epilepsy in Bonn, Germany. Data acquisition was performed by Andrzejak [5]. It consists of five sets A–E which include EEG from a normal person and from those suffering from epilepsy. Three sets were used for our analysis—A, D and E. Set A consists of EEG data from normal people with their eyes closed. Electrodes were placed on the epileptic zone of the brain for Set D and in Set E, data was recorded from individuals who had epilepsy and were having seizures at the time of experiment. The motive behind choosing these sets is that one gives idea about a normal person's EEG when his/her eyes are closed that means there is not much eye activity so EOC/visual artefacts would not be noising the EEG, whereas the other two sets inform about those suffering from epilepsy with or without seizure. Each of these sets consists of 100 trials from a single channel. Data is acquired from A to D convertor with a sampling rate of 173.61 Hz and recorded for a time span of 23.6 s. The 10–20 International Standard of electrode placement is used for data acquisition.

3.2 Wavelet Transforms

Wavelet transforms are very useful for EEG signal analysis because of their nonstationary nature which is similar to that of the signals under study. As stated above wavelet transforms help to decompose the whole signal into wavelet coefficients which then are used for feature extraction. Equation 1 shows the general expression of wavelet transforms, where $x(t)$ is the wavelet, τ stands for the translational motion of the wavelet and s is the scale parameter. $\varphi(t)$ is the transforming function also known as the mother wavelet. Different mother wavelets were used to decompose and reconstruct the signals to check which one gives a waveform closest to the original one. This led the authors conclude that out of all the other wavelet functions like Haar, Symlets, Coiflets, Daubechies (dB) was the most efficient and commonly used one [6]. All these mother wavelets use filters to decompose a signal. Daubechies function varies from dB1 to dB20 based on the order of filter used. As the order increases from 1 to 20, the number of samples chosen also increases which improves the efficiency of reconstruction. Therefore, in this study, Daubechies 20 is used to calculate wavelet coefficients. Figure 1 shows the reconstructed waveform of the delta frequency range using Haar wavelet function, whereas delta, alpha and theta band waveforms are compared in Figs. 2, 3 and 4, respectively. These are reconstructed using dB2, dB8, dB14 and dB20. Undoubtedly, it is evident that Daubechies function gives better results.

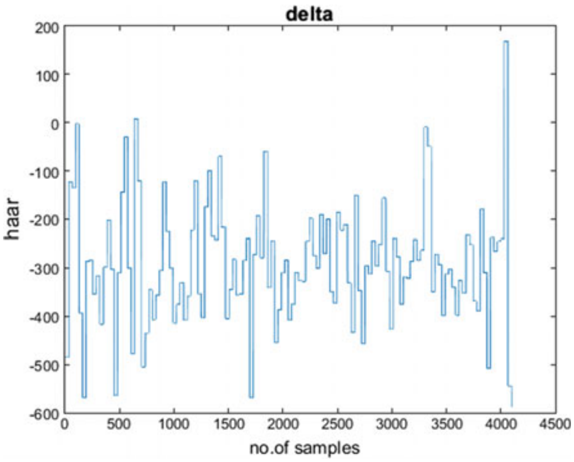


Fig. 1 Delta band plot using Haar wavelets

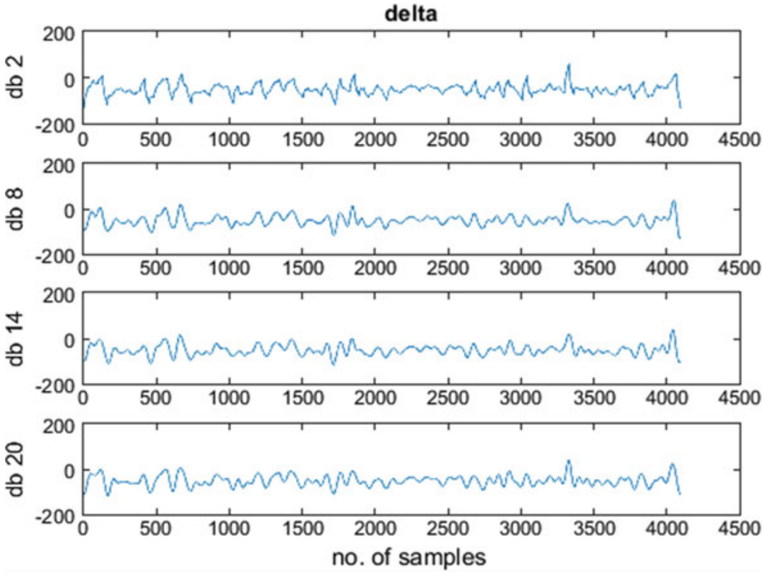


Fig. 2 Delta band plot using different Daubechies family wavelets

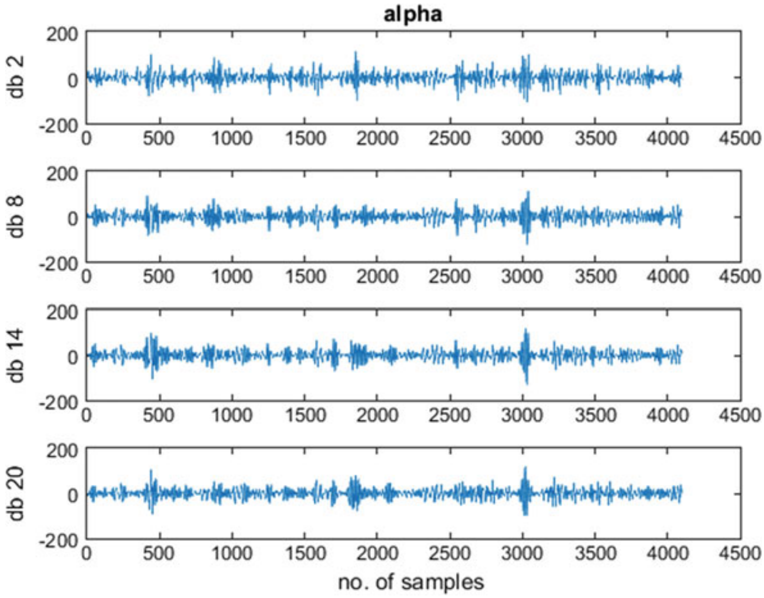


Fig. 3 Alpha band plot using different Daubechies family wavelets

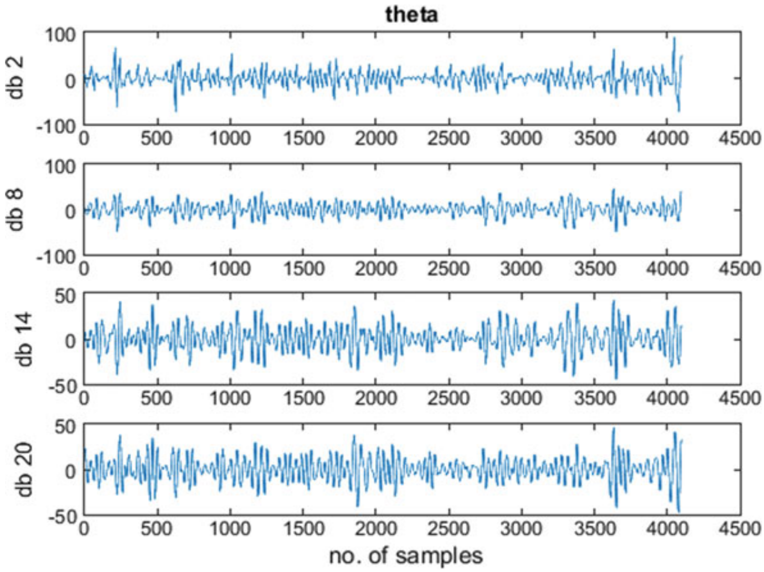


Fig. 4 Theta band plot using different Daubechies family wavelets

$$\varphi(\tau, s) = \frac{1}{\sqrt{|s|}} \int x(t)\varphi^{*}\left(\frac{t-\tau}{s}\right)\mathrm{d}t \quad (1)$$

4 Feature Extraction

The levels of decomposition of an EEG signal from a channel using wavelet transforms is determined by the number of major frequency components that are of interest. Since EEG has its maximum part concentrated from 0 to 30 Hz and for the dataset used, sampling rate is 173.6 Hz, it needs to be divided into five sub-bands. Wavelet decomposition decomposes the signal first into two halves, one is the high-frequency component (using high-pass filter) and other is low-frequency component (using low-pass filter). High-pass filter here follows the wavelet function and low-pass filter is termed as the scaling function [7]. This process needs to be repeated with the lower frequency region till zero is reached. According to the Nyquist criterion, maximum frequency content of the signal is $f_s/2$. To begin with, signal is halved to obtain a lower frequency range from 0 to 86 Hz, half the sampling rate. This is termed as D1 (detailed coefficient level 1). The other half is termed as A1 (approximate coefficient level 1).

Table 1 shows the various frequency ranges in which the signal is divided based on wavelet coefficients. The number of levels of decomposition is calculated based on the sampling rate of the A–D convertor. The frequency range in each level $\left[\frac{f_m}{2} : \mathrm{f}_m\right]$ is related to the sampling frequency and the level of decomposition by the formula mentioned in Eq. 2.

$$f_m = \frac{f_s}{2^{l*l}} \quad (2)$$

where l represents the level of decomposition [8].

Figure 5 shows how a signal is decomposed into three levels using wavelet transforms. Sets A and E were specifically used for wavelet decomposition and feature

Table 1 Wavelet coefficients with corresponding frequency bands and their ranges

Coefficient	Frequency range	Decomposition level/frequency band
D1	43.40–86.80	(1) Noise
D2	21.70–43.40	(2) Gamma
D3	10.85–21.70	(3) Beta
D4	5.42–10.85	(4) Alpha
D5	2.71–5.42	(5) Theta
A5	0.00–2.71	(5) Delta

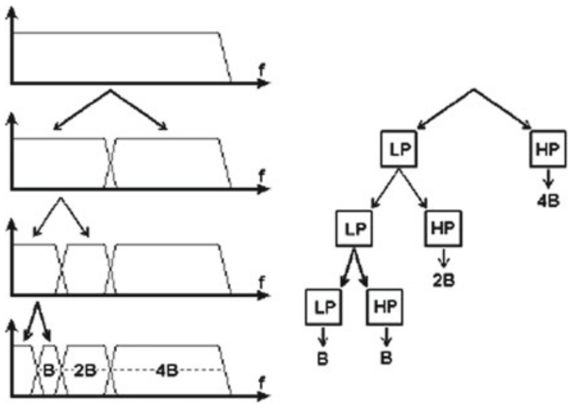


Fig. 5 Wavelet decomposition of a signal into three levels [9]

extraction. The wavelet coefficients from D1 to A5 were used to extract various features as listed below:

1. Maximum of wavelet coefficients in each sub-band.
2. Minimum of wavelet coefficients in each sub-band.
3. Mean of the wavelet coefficients in each sub-band by the formula given as

$$\mu_i = \frac{1}{N}\sum_{i=1}^{N} D_{ij,\ i=1,2,\ldots.l}. \tag{3}$$

4. Standard deviation of coefficients in each sub-band is given as

$$\sigma = \sqrt{\left(\frac{1}{N-1}\sum_{i=1}^{N}\left(D_{ij}-\mu\right)^2\right)}. \tag{4}$$

5. Variance of the coefficients in each band is σ^2.
6. Median of the coefficients of each sub-band is given as

$$\text{Median} = \begin{cases} D_{\frac{(N+1)}{2}}, & \text{if } N \text{ is odd} \\ \frac{1}{2}D\left(\frac{N}{2}\right) + D\left(\frac{N}{2}+1\right), & \text{if } N \text{ is even} \end{cases}. \tag{5}$$

7. Energy of the coefficients of each sub-band: The energy of alpha, beta, etc., bands within an EEG signal were calculated using the formula given below:

$$\text{Energy}(E_i) = \sum_{j=1}^{N}\left|D_{ij}\right|^2, \quad i = 1, 2, \ldots l \tag{6}$$

8. Relative wavelet energy in each sub-band: For this, the energy is first calculated for different wavelet coefficients individually (E_i). Then, the following equation gives the relative wavelet energy of the ith reconstructed coefficient:

$$\rho_i = \frac{E_i}{E_{\text{total}}}. \tag{7}$$

9. Mean square error (MSE): This feature is used to calculate the MSE given as

$$\text{MSE} = \sum_i^N \frac{(D_i - \mu)^2}{N} \tag{8}$$

10. Fractal dimension: It gives idea about the complexity of a signal, how complex are the samples that constitute a signal. If d is the distance between the first sample value and the one which is farthest, fractal dimension is calculated as follows:

$$\text{FD} = \frac{\log_{10} L}{\log_{10} d}. \tag{9}$$

L is the sum of the distance between successive samples. In all the above features, D_{ij} stands for definite wavelet coefficients, N is the total number of wavelet samples and μ is the mean of a particular set of wavelet coefficients. Wavelet coefficients for a normal and epileptic person are plotted in Figs. 6 and 7, respectively. It is evident from the two figures that the amplitude of EEG signals of patients suffering from epilepsy is more than those of normal subjects.

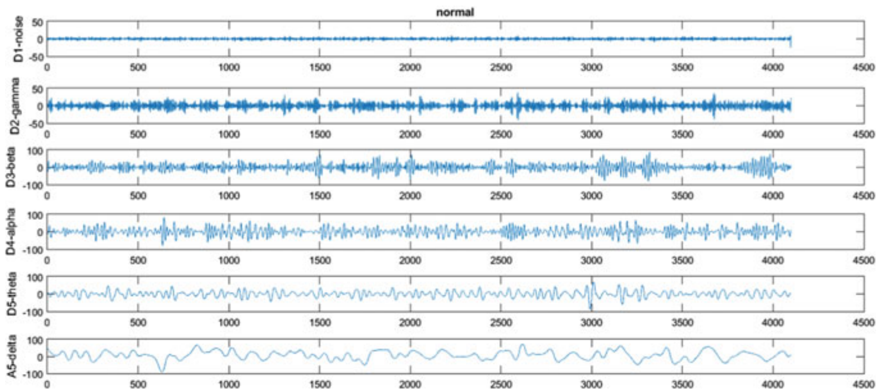


Fig. 6 Wavelet coefficients' plot for a normal person

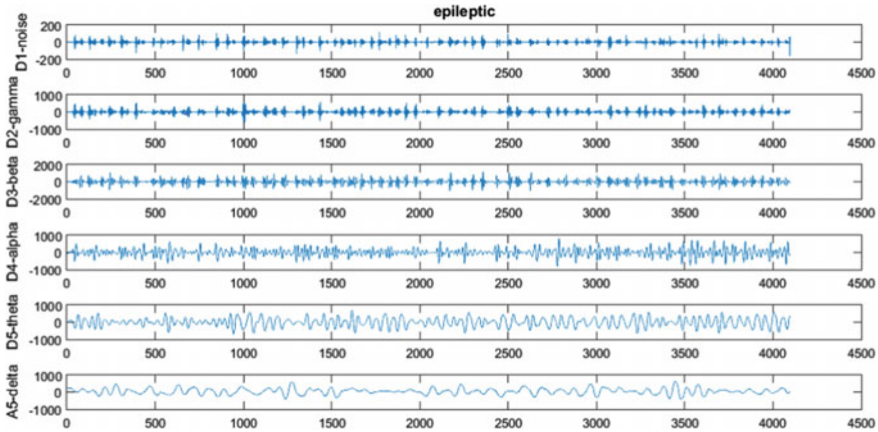


Fig. 7 Wavelet coefficients' plot for an epileptic person

5 Feature Classification

5.1 Artificial Neural Networks (ANN)

Classification can be performed using various tools like support vector machines (SVM), hidden Markov model (HMM), etc. Another efficient tool is the ANN. ANN can be either single layer with no hidden layers or multiple layers where hidden layers are introduced to increase the accuracy of the results obtained. Each neuron in the input layer is connected to all the other neurons in its next layer using some weights or so-called 'biases' [10]. These biases are the result of nonlinear activation functions applied to the input neurons. Number of hidden layers that need to be added and the number of neurons in each layer is determined experimentally and varies according to the application. Most common ANN algorithm in use is the backpropagation algorithm.

5.2 Classification Using ANN

ANN-a multilayer perceptron is used in the paper. These features were fed into a neural network where each input node represents a feature. The model used for this work consists of ten input neurons each representing a feature and two output neurons representing the two classes of subjects—normal and epileptic. In each of these stages, an activation function is used to calculate the final result which can be passed onto the next stage. For the first set of hidden layers, 'tansig' activation function is used, whereas 'purelin' is used for the last set. Features extracted using both normal and epileptic data were randomly distributed to form a dataset. This set

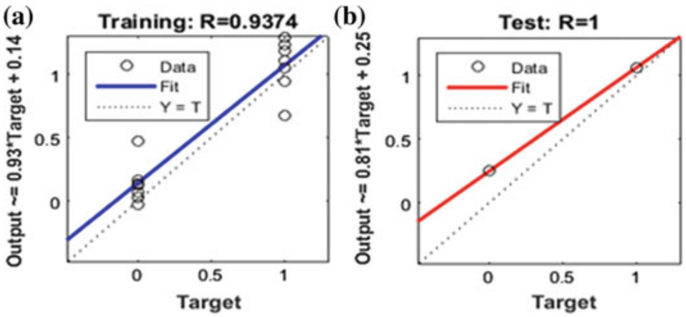


Fig. 8 Regression plot for **a** training data and **b** testing data

was randomly divided into three categories, namely, training, validation and testing. For better analysis, normal subject's dataset was assigned the binary value of 01, whereas that of epileptic patients was assigned 10. After several attempts of training the neural network, the testing dataset was used to test the accuracy of classification. Difference between actual output and that predicted by neural network can be used to calculate MSE. In order to get rid of the redundancies if any, the input dataset used for feature extraction was normalised. To check the maximum accuracy obtained, the network was trained several times using different combinations of hidden layers and neurons in each layer. Regression plots for training and testing data are shown in Fig. 8a, b, respectively. It shows how close our data is compared to the ideal 45° line. Performance of such a system can be evaluated using the validation performance plot shown in Fig. 9a. Performance of the neural network is plotted in terms of MSE. If the MSE converges and is less than one, as is the case here (less than 10^{-2}), it is an indication that the neural network is being trained properly. Network works best at epoch 62 with a gradient of 0.0047866. An epoch gives information about the number of times the training dataset is used to update the weights. Figure 9b shows the confusion matrix obtained using the 15 sets of ten features each. It is evident that out of 15 people only 1 person was erroneously classified. Hence, the overall efficiency obtained is 96.7%. Neural network toolbox in MATLAB is used for the classification.

6 Conclusion

As the number of cases of such neurodegenerative disorders is increasing day by day, urgent attention on research related to their diagnosis is demanded. Early detection can help to lessen if not completely eradicate the symptoms and can make one aware of how to act when onset of seizure takes place. Ten features as stated above are fed to the neural network and after several rounds of training and using hit and trial, the performance was optimised. Minimum MSE obtained had a value less than

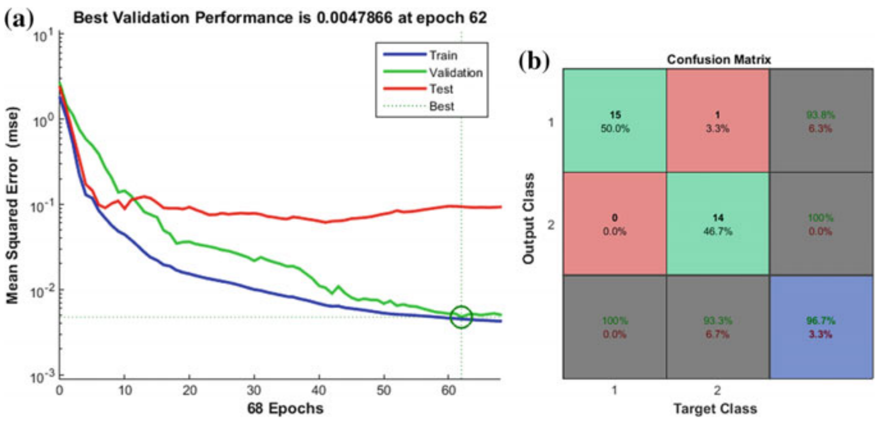


Fig. 9 **a** Performance plot and **b** confusion matrix

10^{-2}. Performance plot and confusion matrix were plotted which gave an accuracy of 96.7%. This work can be extended to develop an application for the detection of epilepsy. The EEG raw data needs to be fed to the application which then can show if the person is suffering from epilepsy or is normal.

Acknowledgements Authors would like to thank the Centre of Epilepsy in Bonn, Germany for the data provided.

References

1. Epilepsy Foundation, http://www.epilepsy.com
2. Hamad A, Houssein EH, Hassanien AE, Fahmy AA (2016) Feature extraction of epilepsy EEG using discrete wavelet transform. In: 12th international computer engineering conference (ICENCO), pp 190–195
3. Kolekar MH (2014) Machine learning approach for epileptic seizure detection using wavelet analysis of EEG signals. In: International conference on medical imaging, m-health and emerging communication systems (MedCom), pp 412–416
4. Guo L, Rivero D, Seoane JA, Pazos A (2009) Classification of EEG signals using relative wavelet energy and artificial neural networks. In: Proceedings of the first ACM/SIGEVO summit on genetic and evolutionary computation, pp 177–184
5. (Dataset) Andrzejak RG, Lehnertz K, Rieke C, Mormann F, David P, Elger CE (2001) Indications of nonlinear deterministic and finite dimensional structures in time series of brain electrical activity: dependence on recording region and brain state. Phys Rev E 64:061907, http://epileptologie-bonn.de/cms/
6. http://wavelets.pybytes.com/
7. Guler I, Ubeyli ED (2007) Multiclass support vector machines for EEG-signals classification. IEEE Trans Inf Technol Biomed xi:117–126

8. Omerhodzic I, Avdakovic S, Nuhanovic A, Dizdarevic Z (2010) Energy distribution of EEG signals: EEG signal wavelet-neural network classifier. World Acad Sci Eng Technol 61:1190–1195
9. Cheong LC, Sudirman R, Hussin SS (2015) Feature extraction of EEG signal using wavelet transform for autism classification. ARPN J Eng Appl Sci x (19):8533–8540
10. Anand SV, Selvakumari RS (2013) Detection of epileptic activity in the human EEG-based wavelet transforms and SVM. Int J Eng Res Technol (IJERT) ii(1):1–6

3D Printed Surgical Guides in Orthognathic Surgery—A Pathway to Positive Surgical Outcomes

Chitra Chakravarthy, Sanjay Sunder, Santosh Kumar Malyala and Ammara Tahmeen

Abstract Rapid advancements in robotics and computer-aided surgeries have revolutionized the field of medicine and surgery. Maxillofacial surgery has benefited from these technological advances as significant contributions have been made in the management of complex soft tissue and bony pathologies. These technologies are especially useful in patients with post-traumatic defects and also cases with esthetic facial deformities. Defects in the craniofacial skeleton are of either congenital, developmental, traumatic, or pathological ethology. The primary purpose of correcting facial anomalies is for functional rehabilitation. The esthetic rehabilitation of a patient is very challenging. It is important to achieve excellent postoperative form and function and also to minimize operative and postoperative morbidity. Rapid prototyping biomodels that are being used in recent times are playing a very significant role not only for patient education, diagnosis of defects but also in surgical planning. Their use in surgical planning reduces anesthesia time, reduces operating time, and provides better esthetic and functional results. The integration of 3D imaging and computerized surgery continues to bring about newer and better changes in the conventional surgeries making the outcome much more beneficial to the patient [1]. We present a case of facial deformity in the form of maxillary excess and retrogenia corrected using orthognathic surgery supported by the use of surgical guides fabricated using additive manufacturing.

C. Chakravarthy · S. Sunder · A. Tahmeen
Department of Oral and Maxillofacial Surgery, Navodaya Dental College and Hospital, Raichur 584103, Karnataka, India

S. K. Malyala (✉)
Department of Mechanical Engineering, National Institute of Technology Warangal, Warangal 506004, Telangana, India
e-mail: msantoshpdd@gmail.com

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_138

1 Introduction

Anterior maxillary osteotomy and genioplasty are some of the most commonly performed orthognathic surgical procedures. Anterior maxillary osteotomy is indicated in patients with vertical or horizontal excess or deficiencies of the anterior maxilla. Genioplasty is a surgical procedure to correct deformity of the chin. Presurgical planning in orthognathic surgery plays an important role in the outcome of the procedure [2]. Conventionally, radiographic analysis and prediction tracings are used to give the surgeon an insight into the area of anomaly and also gave a measure of the extent of the defect in the jaws. Additionally, now 3D facial skeletal analysis has the added advantage of a three-dimensional assessment [3].

The CT scan images are reconstructed into a three-dimensional format on which virtual surgery can be performed and the osteotomy lines accurately planned are taking care of vital structures and getting a clear visualization of the osteotomy on the lingual/palatal aspect which cannot be visualized clinically [3]. The CT scan data can also be converted into a printable format on which a mock surgery can be performed and the osteotomized segment is moved to its new position and the desired result is obtained. This also allows the surgeon to adapt the plates according to the new position of the bone getting the exact contour as will be seen clinically during the surgery. This is extremely advantageous as it saves a lot of time and effort intraoperatively and gives an accurate fit of the fixation plates which is also critical to the outcome of the surgery [4, 5]. The use of CAD/CAM technology can further enhance the predictability of the surgery thereby minimizing the percentage of error. Using the surface data, it is possible to generate various surgical templates or guides which can enable ease of surgery and improved efficiency [6].

The importance of achieving a symmetrical osteotomy cut cannot be overemphasized. Being an esthetic procedure, it is critical to the outcome of the procedure. Using a 3D printed surgical template/guide for an anterior maxillary osteotomy and an advancement genioplasty help us get an accurate osteotomy and reduced surgical time to a large extent. The surgical guide plays an important role in the precise transfer of the 3D surgical planning to the patient in the operating room. This customized printed guide was used to first perform the mock surgery and adapt the plates and then the same was used to guide the osteotomy.

2 Materials and Methods

A 20-year-old female patient presented with a prognathic anterior maxilla and retrogenia. Radiographic analysis and planning were done, presurgical orthodontics was performed for about 8 months prior to surgery after which an orthognathic of anterior maxillary advancement and intrusion and advancement genioplasty were planned.

2.1 Construction of the Surgical Guides

The patient was scanned using a 128 slice SOMATOM perspective CT scanner (Siemens Somata Perspective model number 76970). The scan data was obtained in digital imaging and communications in medicine (DICOM) format which was in two and a half D format. This DICOM data was processed using MIMICS 18.0. materialized medical software to generate a 3D CAD model of the patient's scanned data. This data completely included the hard and soft tissue contents. By adjusting the threshold value, the bone data was isolated. Using MIMICS software a threshold value of 226–3071 Hounsfield units can be applied to get accurate bone detail.

The next step was to separate the region of interest from the complete scan. The mandible was first separated and a 3D CAD model was created. Similarly, the maxilla up to the infraorbital margins was isolated creating a separate 3D CAD image. These were then converted into stereolithography (STL) file formats. This is a neutral file format for all AM machine preprocessing software. The 3D medical models of maxilla and mandible were fabricated using material extrusion technique.

Virtually, a surgical simulation of the surgery was performed on the maxilla and mandible. The genioplasty cut was simulated taking care of the mental nerves, the root apices and also inclining the cut as required from the buccal to lingual cortex. The osteotomized segment was also moved virtually to present an ideal three-dimensional picture of the outcome of the surgery. The maxillary anterior segmental osteotomy was also simulated virtually keeping the cuts exactly 5 mm apart as planned.

Using the surface data generated on the 3D CAD model, a surgical template was designed (Fig. 1a) and various designs of genioplasty guides have been reported till date [7]. This genioplasty template was designed to extend from the mandibular anterior teeth all the way to the lower border of the mandible taking care to isolate the mental nerves bilaterally. The purpose of the lower border extension was to improve the accuracy of the fit of the template and also to ensure the osteotomy cuts extend into the lower border. A slot was created exactly along the virtual osteotomy line. The edges of the slot were angulated at a 30° chamfer to enable the drill or saw to follow that angulation (Fig. 1b). The maxillary template was extended along the lateral border of the piriform aperture till the second premolar on both sides. A 5 mm slot was created to guide the saw for the two cuts (Fig. 1c). The maxillary surgical guide followed the border of the piriform aperture. These surgical templates followed the surface data accurately to get a perfect adaptation. These surgical templates were then printed using a biocompatible material.

2.2 Mock Surgery

Using the templates and the 3D printed jaws, the mock surgery was performed presurgically and the virtual cuts were replicated on the models using the template. The surgical guide was then removed and the osteotomized segments were separated

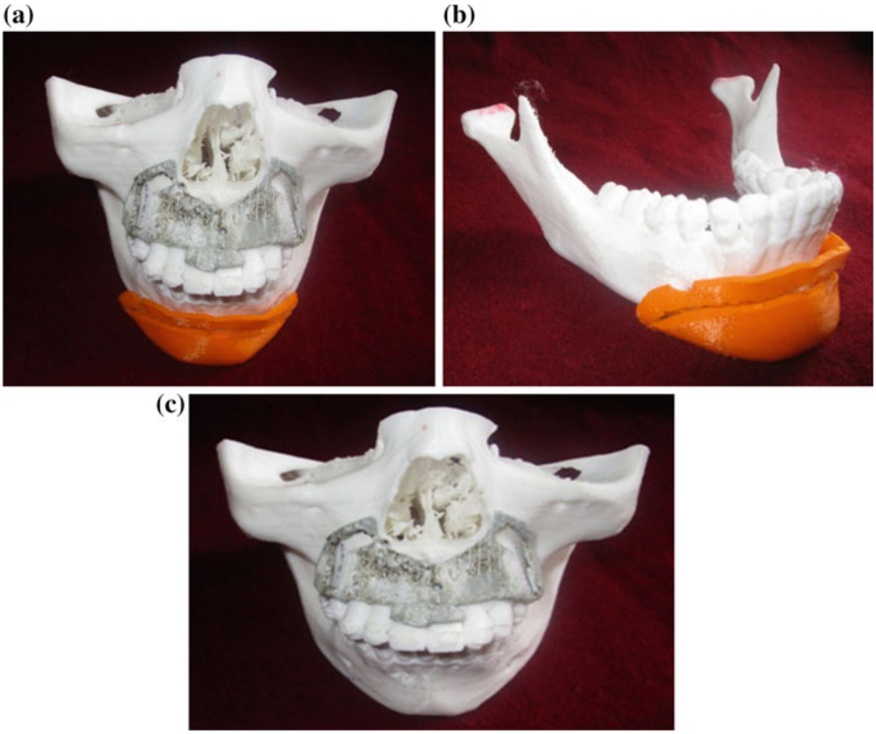


Fig. 1 **a** Mandible and maxillary templates, **b** Slot extensions on the genioplasty template, **c** Slot extension on the maxillary template

and repositioned as planned in the virtual surgery (Fig. 2a, b). Titanium plates were then meticulously adapted to follow the bony contours of the maxilla in its new position. The genial segment was also advanced and titanium plates were adapted to the advanced segment (Fig. 3a, b). These same pre-bent/adapted plates were then to be used during surgery after being sterilized. The surgical guides were also sterilized and used during surgery.

2.3 Surgical Procedure

Following the mock surgery, an informed consent was obtained from the patient. Under general anesthesia, the anterior mandible was exposed to the lower border. The mental nerves were identified and isolated bilaterally. The template was placed and fixed using a single monocortical screw (Fig. 4a). A surgical saw was used to perform the osteotomy as was planned following the chamfer line. The cut was extended to the lingual cortex and to the lower border. With the use of the template, the

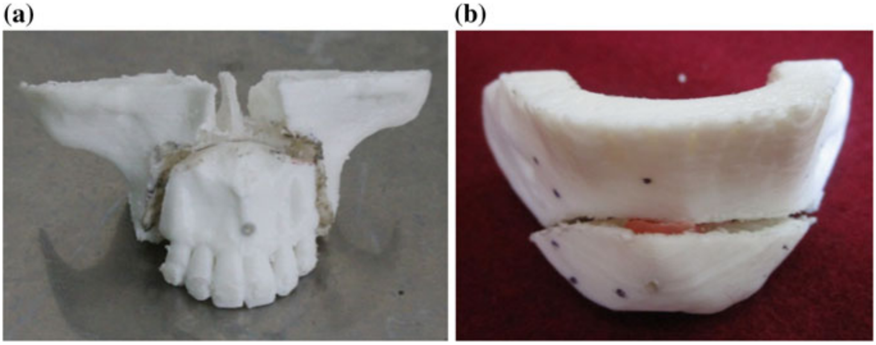


Fig. 2 **a** Osteotomized segments of maxilla, **b** Osteotomized segments of mandible

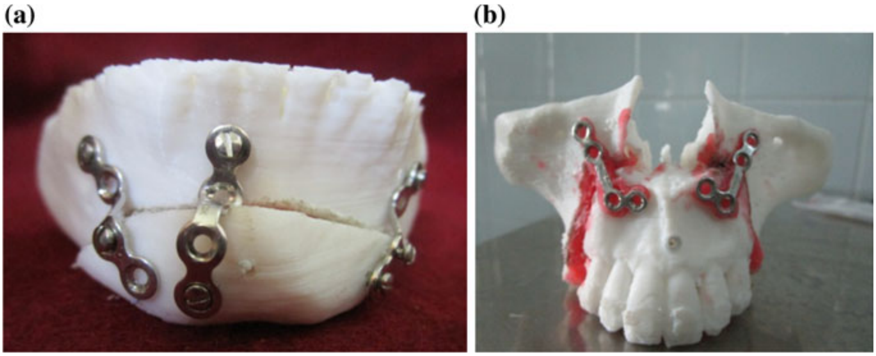


Fig. 3 **a** Adaptation of plates on 3D model of mandible, **b** Adaptation of plates on 3D model of maxilla

symmetry and position of the cut were accurate. The time taken for the surgery was also significantly reduced. Similarly, the anterior maxillary osteotomy was carried out using the templates and setback was done (Fig. 4b). The templates were removed and the osteotomized segments were mobilized to their desired positions. The positions were determined by the adaptation of the plates. Since the plates were already pre-contoured, the surgical time was further reduced. The plates were fixed as planned in the mock surgery.

3 Results and Discussion

Using the 3D printed surgical guides during osteotomy significantly reduced the time of surgery and also the osteotomy cuts could be placed more precisely without any risk of overextension or asymmetrical cuts. Postoperative results were satisfactory and the symmetry could be appreciated. Orthognathic surgery is an elective cosmetic

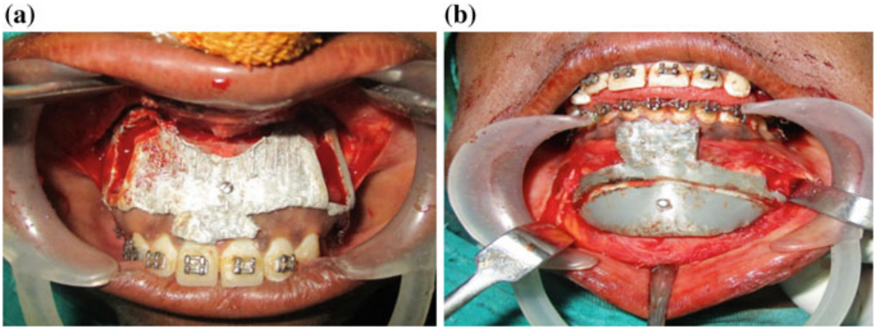


Fig. 4 **a** 3D template placed on the maxilla using a monocortical screw, **b** 3D template placed on the mandible using a monocortical screw

procedure. The outcomes of all the surgery depend to a large extent on the accuracy of the osteotomies. Symmetrical osteotomy and position of the cuts make the surgery successful. The use of the surgical template enables us to visualize the surgery virtually to make changes if necessary and also to replicate the same osteotomy line in the patient intraoperatively. The use of customized wafers and pre-bent plates has been gaining importance in orthognathic surgery [8].

Fabrication of the surgical guide has to be done with great planning keeping in mind certain key points in the design. The contour of the template has to accurately match and fit the contour of the jaw. This is of utmost importance because we do not want it to move during the osteotomy. The thickness of the template should be adequate to ensure that the template does not bend or deform during the surgery. If it is designed too thick or overextended, it will interfere with vital structures in the area. Bulky templates are difficult to use especially in the maxillofacial region where the access is limited. We should also make sure that the template is not cut during the surgical procedure which may possibly put traces of the template material into the surgical site. In this particular genioplasty template, 30° chamfer was very useful as it allows the advancement genioplasty osteotomy to be performed accurately. With the help of the 3D printed models, the patient and the clinician benefit in different ways. The models help in an accurate diagnosis and give an assessment of the extent of the facial deformity. This can be demonstrated to the patient for easy understanding of the defect. It also improves patient compliance as they are able to visualize the problem area and understand the way it needs to be treated. Surgical models help the practitioner in performing the mock surgery and visualizing the desired outcome of the surgery before it is actually replicated on the patient. This helps in improving the confidence of the surgeon. Fabrication of surgical templates is one of the major advantages of 3D printing. Once the mock surgery is done, a precise transfer of the same surgical planning can be done to the operating table with the help of these surgical guides. It helps us to replicate the mock surgery.

Adaptation of the titanium plates on the 3D model has several advantages. A lot of time and effort goes into adaptation of the plates during the surgery. Adding to the difficulty is also the presence of blood and soft tissue obstructing the view of the bony contours. Preadapted plates significantly reduce the surgical time, the exposure time to general anesthesia. It also decreases the total amount of blood loss and the time that the wound is exposed to the external environment.

The preadapted plates serve another important function. They serve as a guide for the mobilization of the osteotomized segment. This is especially useful in cases of advancement genioplasty where the segment has to be advanced arbitrarily. The bent plates serve as a guide to the amount of advancement required. The plates can be fixed accurately without wasting much operating time. Conventionally, the maxillary osteotomy takes about 90 min in our center of which significant time is taken for assessing the symmetry and for contouring the plates both of which were done easily. The surgical time was reduced to an hour for the maxillary procedure. The genioplasty also usually takes about an hour which was completed using the surgical guides in about 40 min. Fabricating of a 3D printed model or guide is now a relatively simple process for a clinician. An accurate CT scan with 1 mm cuts helps to replicate an accurate image. A helical scanner with a zero-degree gantry tilt scanning parallel to the occlusal plane is preferable to reproduce the jaws accurately. The DICOM images are converted into STL images in the laboratory. Using the MIMICS software, specific virtual cuts can be made on the bone and the osteotomized segments can be moved as desired. Some of the challenges we face using this technology need to be overcome. Since this is a relatively new field, changes are still in progress to achieve maximum benefit.

The accuracy of the models, especially in areas where the bone is thin may not be perfect. A major part of maxillofacial surgeries such as trauma and orthognathic are dependent on achieving an ideal occlusion, replication of the precise contours of the dental surface is of prime importance. This is not always possible with even a high-resolution CT scan. Sometimes a part of the data is lost due to the presence of metal artifacts such as crown and bridges, and orthodontic brackets. To overcome this issue, dental casts of the patient are scanned separately using a high-resolution scanner and the teeth are superimposed to recreate the delicate details of the teeth in the jaw. These minor drawbacks do not pose as an obstacle in the path of this evolving field as the advantages far outweigh the drawbacks.

4 Conclusion

Orthognathic surgery is a procedure that is done to improve facial esthetics. Symmetrical and accurate osteotomies and precise mobilization of the segments is of prime importance to the outcome of the surgery. The use of modern day technology in the form of additive manufacturing has added precision and ease to this challenging procedure. The use of a surgical guide allows the surgeon to easily perform the surgery as planned in the mock surgery and in the virtual surgery. It simplifies the surgery and allows improved accuracy and superior surgical outcomes. Adaptation of this technology into other surgical aspects of maxillofacial surgery will definitely give an added benefit to the surgeon in the form of reduced laboratory time used for planning [9], reduced surgical time, ease of surgery, and improved clinical outcomes.

Acknowledgements The authors declare that they have no conflict of interest.

References

1. Xiao Y, Sun X, Wang L, Zhang Y, Chen K, Guomin W (2017) The application of 3D printing technology for simultaneous orthognathic surgery and mandibular contour osteoplasty in the treatment of craniofacial deformities. Aesthetic Plast Surg 41(6):1413–1424
2. Haas OL Jr, Becker OE, de Oliveira RB (2014) Computer aided planning in orthognathis surgery—systematic review. Int J Oral Maxillofac Surg, pii: S0901-5027(14)00430-5
3. Wang L, Tian D, Sun X, Xiao Y, Chen L, Wu G (2017) The precise repositioning instrument for genioplasty and a three-dimensional printing technique for treatment of complex facial asymmetry. Aesthic Plast Surg 41:919–929
4. Rubio-Palau J, Prieto-Gundin A, Cazalla AA, Serrano MB, Fructuoso GG, Ferrandis FP et al (2016) Three-dimensional planning in craniomaxillofacial surgery. Ann Maxillofac Surg 6:281–286
5. Raphaël O (2013) Three-dimensional rapid prototyping models in cranio-maxillofacial surgery: systematic review and new clinical applications. P Belg Roy Acad Med 2:43–77
6. Lim S, Kim M, Kang S (2015) Genioplasty using a simple CAD/CAM (computer-aided design and computer aided manufacturing) surgical guide. Maxillofac Plast Reconstr Surg 37:44
7. Choi JW, Namkug K (2015) Clinical application of three-Dimensional printing technology in craniofacial plastic surgery. Arch Plast Surg 42:267–277
8. Suojanen Juho, Leikola Junnu, Stoor Patricia (2016) The use of patient = specific implants in orthognathic surgery: a series of 32 maxillary osteotomy patients. J Cranio-Maxillofac Surg 44:1913–1916
9. Wrzosek MK, Peacock ZS, Laviv A, Goldwaser BR, Ortiz R, Resnick CM, Troulis MJ, Kaban LB (2016) Comparison of time required for traditional versus virtual orthognathic surgery treatment planning. Int J Oral Maxillofac Surg 45(9):1065–1069

Image-Based Method for Analysis of Root Canal Geometry

Ankit Nayak, Prashant K. Jain, P. K. Kankar and Niharika Jain

Abstract Image-based techniques of observation and measurement are getting popular due to its accuracy and accessibility. In dentistry, endodontic files are used for root canal shaping and cleaning. Manufacturers design and fabricate different endodontic files. Endodontists examine the performance of different endodontic files. Endodontists used artificial root canal for comparison and examination of the file system. After shaping of artificial root canals, endodontic blocks are compared by qualitative observation. This article presents a methodology based on image analysis of artificial root canal for examination and comparing the shaping ability of endodontic files. In the proposed method, the images of the artificial root canal were processed in a custom-made MATLAB® program to quantify the root canal deviation and root canal transportation. Calculation of root canal curvature quantified root canal transportation. Performance of 20/0.02 NiTi hand file and 20/0.04 NiTi rotary file for glide path preparation was compared. Twenty endodontic blocks were divided into two experimental groups for canal preparation. Pre- and post-treatment root canals were compared regarding canal curvature modification and deviation for the experimental group of 20/0.02 NiTi hand file with self-adjusting file (SAF) and 20/0.04 NiTi rotary file with SAF using *t*-test at 95% confidence interval. Pre- and post-treatment images are processed in the custom-made MATLAB® program to find out the canal

A. Nayak (✉) · P. K. Jain · P. K. Kankar
CAD/CAM Lab, Mechanical Engineering Discipline,
PDPM Indian Institute of Information Technology, Design and Manufacturing,
Jabalpur 482005, Madhya Pradesh, India
e-mail: ankitnayak@iiitdmj.ac.in

P. K. Jain
e-mail: pkjain@iiitdmj.ac.in

P. K. Kankar
e-mail: kankar@iiitdmj.ac.in

N. Jain
Departments of Conservative Dentistry and Endodontics,
Triveni Institute of Dental Science, Hospital and Research Centre,
Bilaspur 495001, Chhattisgarh, India
e-mail: niharika.dr@gmail.com

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_139

transportation and deviation. A significant difference in an experimental group of hand file and the rotary file was found. It has been found that canal transportation in case of 20/0.02 NiTi hand file was significantly greater than 20/0.04 NiTi rotary file.

1 Introduction

Root canal shaping is the subject of excellent care and attentiveness. Root canal treatment (RCT) is the favored treatment process for the infected tooth. Endodontists used endodontic files to remove the infected tissues from the root canal followed by root canal filling and sealing. Root canal shaping is an essential step of RCT, which should be performed with great care, the success of RCT depends on the root canal filling. First, root canal shaping was done with the help of watch spring by Edwin Maynard of Washington, D.C. in 1838 [1]. After that several types of endodontic instruments were designed and fabricated by manufacturers. These instruments are based on different mechanics of root canal shaping, shaping kinematics, and manufacturer's protocols. Endodontic instrument helps dentists up to the great extent to achieve accuracy in root canal shaping. Axis deviation or root canal transportation [2–6] is the cause of root perforation [7], external tooth resorption, and tooth fracture [8, 9]. For examination and comparison of the performance of endodontic files, endodontists are using different image-based methods like X-ray, photography, or computed tomography (CT) scanning [10, 11].

CT scanning is a costly means to know the geometry of root canal, while in case of radiography, there is not any established method to extract the information from the radiograph of teeth. Some researchers have used photography for evaluation of canal curvature of endodontic blocks (Endo Training Bloc-J, Dentsply Maillefer, Ballaigues, Switzerland) [12]. Endodontic blocks are getting famous for evaluation and comparisons of different endodontic file systems [13, 14]. Since material properties and architecture of root canal are uniform for each endodontic block, this helps to maintain the uniformity in the sample population, while it is not easy to keep in case of extracted teeth. Moreover, extracted teeth of the same anatomy and uniform architecture of root canal are not natural to get for experimentation.

In this article, endodontic training blocks are used for experimentation to compare the shaping ability of the SAF [12], while it used with 20/0.02 (hand) and 20/0.04 (rotary) files for glide path preparation. After the glide path preparation, SAF is used for shaping the root canal. Prepared root canals were examined using custom-made MATLAB program for image analysis of post- and pre-treated images of endodontic blocks to find out the deviation in root canal axis.

2 Material and Method

Twenty J-shaped endodontic blocks are used for ex vitro study of endodontic files. The study is divided into three steps pre-processing, root canal shaping, and post-processing. Pre-processing of endodontic block involves cleaning of endodontic blocks in order to remove dirt and stain and this helps to get error-free images of root canal; after that the root canals are filled with red ink so that it can be identified in pre-processing images. In root canal shaping, endodontic files are used for shaping of simulated root canal as per the file manufacturer protocol for the respective endodontic file. After root canal shaping, the simulated root canals are filled with black ink and scanned again for post-processing images. Pre- and post-instrumented images were processed in MATLAB to remove noise and image artifact. Median filter by taking 2×2 sliding window was applied to remove the salt pepper noise from images. An optimum contrast value for each image was used to eliminate the image artifact. Pre-processed images were sent to the MATLAB code for analysis. Image analysis algorithm has been explained in the following flowchart.

Centering ability of root canal can be analyzed by measuring the deviation between the axis of pre- and post-operated root canals. Pre- and post-treatment images of root canal have been taken to compare and collect the information of trajectory of root canal shaping. Pre- and post-treatment images are converted into binary images followed by thinning of root canal up to one-pixel width to get the central axis of root canals. Geometric analysis of root canal axis can collect information regarding canal modification. Canal transportation and curvature modification were analyzed using custom-made MATLAB programs (Fig. 1).

Coordinates of root canal axis were calculated from the binary image of root canal axis. The axis of root canal has been converted into the series of point coordinates. Then, a circle was fitted into the curvature of root canal using Eq. 1.

$$(x - x_c)^2 + (y - y_c)^2 = r^2 \tag{1}$$

where $x_c y_c$ are the coordinates of the center of circle and r is the radius of the circle.

Radius and the center of curvature for root canal were calculated. The curvature of original and modified root canal has been compared to know about the changes in root canal geometry (Fig. 2).

Images, showing the center axis of the root canal, are superimposed and the region of interest (ROI) is cropped. Superimposed image shows the axes of pre- and post-operated root canals. After the Boolean operation between cropped and superimposed image, the flood-fill algorithm was applied. Flood-fill algorithm helps to remove the noise that occurs after the skeletization of the root canal image. This image will contain the boundary of white pixels which are not the part of deviation. To eliminate these pixels, the superimposed image was subtracted from the image of deviation. After that the pixels were counted to get the deviation in any particular case of the instrumented root canal (Fig. 3).

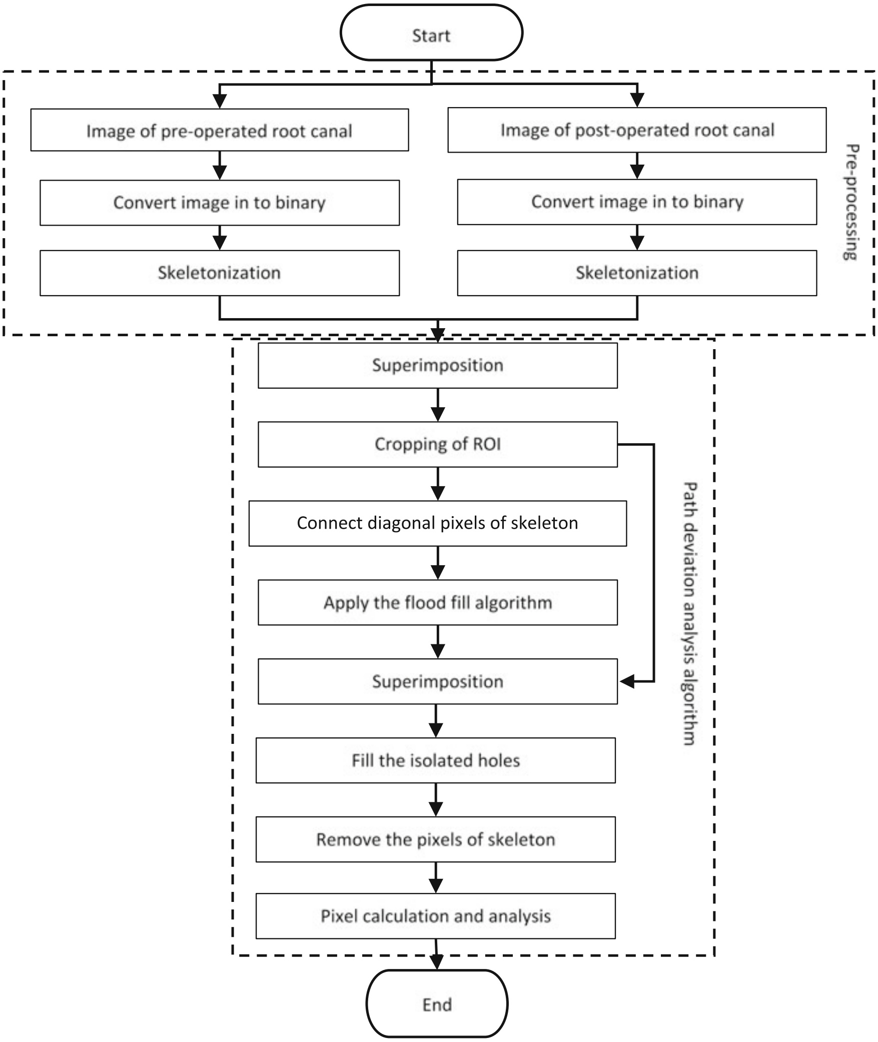


Fig. 1 Flowchart of the image analysis algorithm

3 Results

Post- and pre-treated images for hand file and mechanized file are processed through the proposed algorithm to analyze canal modification in terms of root canal transportation. The results of canal curvature modification and deviation for the experimental group of 20/0.02 NiTi hand file with SAF ($n = 10$) and 20/0.04 NiTi rotary file with SAF ($n = 10$) are compared using t-test at 95% confidence interval. Results of t-test are shown in Table 1.

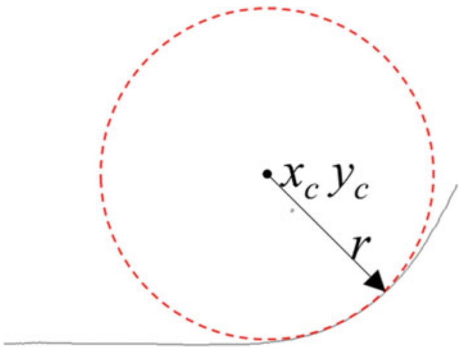


Fig. 2 Fitted circle in the root canal curvature

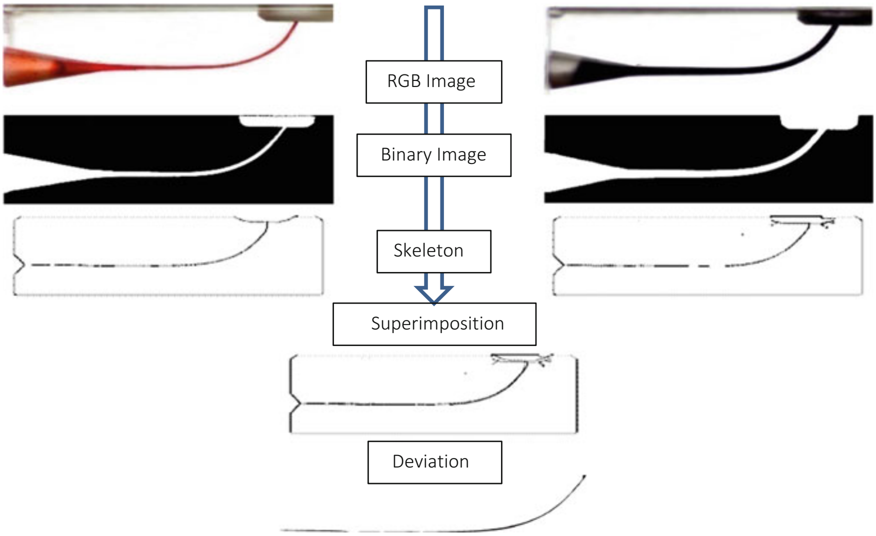


Fig. 3 Major steps of image processing

Table 1 Deviation in root canal curvature of the experimental group (Mean ± Sandard deviation)

Parameter	Group A	Group B	*P* value
Root canal deviation	14,783 ± 1182	7548 ± 648	$P < 0.05$
Root canal curvature modification	343 ± 24	295 ± 20	$P < 0.05$

P value for two-sample *t*-test at 95% CI is less than 0.05, which means that there is a significant difference in root canal curvature modification and root canal deviation between group A and group B.

4 Discussion

The image-analysis-based algorithm is developed to compare the performance of the SAF in endo training blocks, while glide path was prepared using 20/0.02 (hand) and 20/0.04 (rotary), two different files. Glide path preparation, coronal enlargement is essential for the good results of RCT. Endodontists have used a different kind of file for glide path and root canal preparation [14, 15]. Different root canal and glide path preparation techniques and protocols are introduced and tested by the endodontists. These techniques are unable to compare or examine the approaches quantitatively [3, 16, 17]. The suggested method of image analysis can be used for comparison of various processes, endodontic files, and protocols of RCT.

Computed tomography (CT) scan gives three-dimensional details of root canal architecture. CT scan is used for detailed analysis of prepared root canal [11, 18]. CT process is much costly than image scanning, and post-processing of CT data is also complicated to collect the information of root canal geometry. Instead of CT scanning, suggested methodology of image analysis is less costly to examine the endodontic block. However, it is less accurate than CT analysis but it is under the limit of tolerance zone.

Acknowledgements Endodontic blocks which are also known as J-shaped resin blocks manufactured by Dentsply Maillefer, Ballaigues, Switzerland have been used for experimentation.

References

1. Grossman LI (1982) A brief history of endodontics. J Endod 8:2–5
2. Goldberg F, Massone EJ (2002) Patency file and apical transportation: an in vitro study. J Endod 28:510–511
3. Bürklein S, Jäger PG, Schäfer E (2016) Apical transportation and canal straightening with different continuously tapered rotary file systems in severely curved root canals: F6 SkyTaper and OneShape versus Mtwo. Int Endod J 50:983–990
4. Zanesco C, Só MVR, Schmidt S, Fontanella VRC, Grazziotin-Soares R, Barletta FB (2017) Apical transportation, centering ratio, and volume increase after manual, rotary, and reciprocating instrumentation in curved root canals: analysis by micro-computed tomographic and digital subtraction radiography. J Endod 43:486–490
5. González Sánchez JA, Duran-Sindreu F, de Noé S, Mercadé M, Roig M (2012) Centring ability and apical transportation after over instrumentation with ProTaper universal and ProFile Vortex instruments. Int Endod J 45:542–551
6. Nayak A, Jain PK, Kankar PK, Jain N (2018) Computer-aided design–based guided endodontic: a novel approach for root canal access cavity preparation. Proc IMechE Part H: J Engg Medicine 232:787–795
7. Tsesis I, Fuss Z (2006) Diagnosis and treatment of accidental root perforations. Endod Top 13:95–107
8. Cheng R, Zhou X-D, Liu Z, Hu T (2007) Development of a finite element analysis model with curved canal and stress analysis. J Endod 33:727–731
9. Uddanwadiker RV, Padole PM, Arya H (2007) Effect of variation of root post in different layers of tooth: linear vs nonlinear finite element stress analysis. J Biosci Bioeng 104:363–370

10. Gergi R, Rjeily JA, Sader J, Naaman A (2010) Comparison of canal transportation and centering ability of twisted files, pathfile-ProTaper system, and stainless steel hand K-files by using computed tomography. J Endod 36:904–907
11. Gergi R, Osta N, Bourbouze G, Zgheib C, Arbab-Chirani R, Naaman A (2015) Effects of three nickel titanium instrument systems on root canal geometry assessed by micro-computed tomography. Int Endod J 48:162–170
12. Nayak A, Kankar PK, Jain N, Jain PK (2018) Force and vibration correlation analysis in the self-adjusting file during root canal shaping: an in-vitro study. J Dent Sci 13:184–189
13. Schrader C, Peters OA (2005) Analysis of torque and force with differently tapered rotary endodontic instruments in vitro. J Endod 31:120–123
14. Choi DM, Kim JW, Park SH, Cho KM, Kwak SW, Kim HC (2017) Vibrations generated by several nickel-titanium endodontic file systems during canal shaping in an ex vivo model. J Endod 1–4
15. Hof R, Perevalov V, Eltanani M, Zary R, Metzger Z (2010) The self-adjusting file (SAF). Part 2: mechanical analysis. J Endod 36:691–696
16. Jain N, Pawar AM, Ukey PD, Jain PK, Thakur B, Gupta A (2017) Preservation of root canal anatomy using self-adjusting file instrumentation with glide path files, pp 51–55
17. Capar ID, Ertas H, Ok E, Arslan H, Ertas ET (2014) Comparative study of different novel nickel-titanium rotary systems for root canal preparation in severely curved root canals. J Endod 40:852–856
18. Benyo Szilagyi L, Haidegger T, Kovacs L, Nagy-Dobo CB (2009) Detection of the root canal's centerline from dental micro-CT records. In: 31st annual international conference IEEE EMBS, pp 3517–3520

Analysis of Explicit Parallelism of Image Preprocessing Algorithms—A Case Study

S. Raguvir and D. Radha

Abstract The need for the image processing algorithm is inevitable in the present era as every field involves the use of images and videos. The performance of such algorithms can be improved using parallelizing the tasks in the algorithm. There are different ways of parallelizing the algorithm like explicit parallelism, implicit parallelism, and distributed parallelism. The proposed work shows the analysis of the performance of the explicit parallelism of an image enhancement algorithm named median filtering in a multicore system. The implementation of explicit parallelism is done using MATLAB. The performance analysis is based on primary measures like speedup time and efficiency.

1 Introduction

Parallel computing is a type of computation in which many calculations are carried out simultaneously, operating on the principle that large problems can often be divided into smaller ones, which are then solved at the same time. Parallel computations can be carried out in different ways like bit-level computations, instruction-level computations, and data- and task-level computations. Even though parallelism has been thought and employed for many years, it could not be used widely because of physical constraints; however, parallel computing has become the main requirement in computer architecture, mainly in the form of multicore processors.

Parallelism of a task in computers depends on the hardware having multicores and multiprocessors with multiple processing elements within a machine. This is different from performing the same task in multiple computers as in grids, multiple passive arrays (MPP), and clusters. There are some specialized parallel computer architectures that are used beside traditional processors, for speeding up the specific tasks.

S. Raguvir · D. Radha (✉)
Department of Computer Science & Engineering,
Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Bangalore, India
e-mail: dm_radha@yahoo.com

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_140

Applications are often classified depending on the way subtasks are required to get communicated or synchronized with each other. An application needs complex parallelism if the divided tasks have to communicate with each other multiple times per unit time. In another way, applications show simple parallelism if there are subtle communications between the subtasks. It exhibits embarrassing parallelism if subtasks rarely or never have to communicate.

When implicit parallelism is employed in MATLAB, but there are some restrictions on the code where we cannot add parallelism into it. Even though there are different ways of doing the parallelism, there are some overheads behind them to get the synchronized output.

2 Literature Survey

The parallelism can be thought of parallel computing and distributed computing. In parallel computing, the parallelism is done in one/more CPUs in the same system, whereas in distributed computing it happens with multiple processors connected by communication links. The most important reason to prefer parallel computing over traditional sequential computing is the ease in applying algorithms, reduced complexity, and increased standardization [1].

Parallel computing can be in the form of implicit and explicit parallelism. Certain programming languages use automatic parallelism available in it. They are named as implicit parallelism. Explicit parallelism can be achieved by parallel computations using special-purpose directives or function calls available in the programming languages. Most of the applications require subtasks like synchronization, communication, or task partitioning to achieve parallelism. Apart from contributing to the intended parallelism of the program, there are overheads which are required to maintain the synchronization.

Various applications or fields need parallel computing for balancing the requirement of the current world. A massive task with fast execution is the requirement of the competitive world. Medical imaging, network-based processing, routing, bioinformatics, and data analytics are fields to name a few. Image processing is vital in computer vision, medical imaging, meteorology, astronomy, remote sensing, and other related fields [2]. Images of any size consume more time to do any processing. Parallel computing can be an apt way to utilize the essence of image processing applications. Distributing visualization algorithms has shown increased efficiency and increased speedup with respect to the processors used [3].

Image processing involves large quantities of data and certain applications use the same kind of operations that are performed on every part of the image. Parallelism applied to such applications is an advantage. It can be applied in various ways like data parallel, task parallel, and pipeline parallel. In data-parallel approach, each computing unit will get a piece of data and perform the operations on receiving data. The primary challenge of this approach is in making the parallelism efficiently in decomposing of data and composing the data back to its form. The next challenge is to balance the load. The image data should be distributed in such a way that the computing units get approximately the same load. This avoids unnecessary communication between computing units. In task-parallel approach [4], certain operations are grouped in tasks and will be allotted to computing units. The approach also requires efficient planning of selecting and distributing the tasks to various computing units. In pipeline-parallel approach, parallelism can be implemented if image processing application requires multiple images so that images are processed in different stages at the same time.

Parallel computing using MATLAB is one of the most widely used and easy to learn platforms. It holds many toolboxes for various applications which made it easy to use [5] and as it follows most of the basic programming languages. It is a platform for learning mathematical computing environments in technical computing. In the paper [6], authors have discussed the use of MATLAB in parallel computations of various research fields such as signal and image processing. It also says about the parallel computing toolbox (PCT) used for parallel computations.

MATLAB can utilize the multicore processors using workers that run locally [7]. The work discussed the fast and efficient computation in multicore processors. There are different ways of achieving parallelism in MATLAB. MATLABMPI, bcMPI, pMATLAB, Star-P, and PCT are to name a few. And, also, the future of MATLAB for parallel computations was discussed.

The advantages of the PCT are that it supports high-level constructs, for instance, parallel for loops and distributed arrays, it has various math functions which let the users utilize existing MATLAB. The simplest way of doing parallelism in MATLAB is using parfor, spmd-end constructs. Parfor and spmd-end assert that all the iterations of the loop are independent so that it can be done in any order or in parallel. It is easy to combine serial and parallel codes in parfor and spmd. There will be better speedup with little effort. Identification of independence in the algorithm is a trade-off for the speedup with the least effort. Independence can be achieved through interprocess communication [8]. But, such conditions can be avoided in explicit parallelism and lead to distributing the load among different processors/cores/threads by labindex of the MATLAB. Each labindex is associated with a core which improves the performance and efficiency of the system.

These parallel algorithms can work with various numbers of threads, to take all the benefits of the upcoming processors having any number of cores.

Parallelization is an add-on to the computing environment. The images are processed and analyzed to find the accuracy and reliability of the results of parallelism. The processing time is also to be considered along with its accuracy and reliability. More the processing time in parallelization, degrades the use of parallelism in the applications. Like the other stages of image processing such as preprocessing used for clarity and visibility of the image, parallelization involves in the optimization of the speed at which the image is processed.

The rapid requirement of parallelism increased the attention to use polyhedral frameworks for optimizing explicit parallelism [9].

The proposed work explores current multicore architectures to parallelize the image processing algorithms which enhance the image. Comparisons on sequential and parallelization through the usage of different number of cores available are discussed.

These parallel algorithms can work with a different number of threads and their processors with different number of cores. Load distribution among processors can be of different types like horizontal, vertical, and equal distribution (cubical).

Image processing algorithms can be parallelized using the simplest way by distributing the load to different cores than considering the limitations of many other aspects of different ways of parallelizing the algorithms.

Parallel computing is having very important significance in several image processing techniques like edge detection, histogram equalization, noise removal, image registration, image segmentation, feature extraction, different optimization techniques, and many more [7].

A different approach of parallel programming from the programmer's point of view is discussed which is known as explicit parallelism [2]. Comparison of sequential and parallel version and the performance of them are discussed.

Even though there is no full acceptance on explicit parallelism.

3 Proposed Model and Implementation

Parallelism in image processing is essential in every field for the current world. There are different ways of parallelizing the algorithms for an image. Every method is having its own advantages and disadvantages.

To avoid the limitations of various parallel programming constructs, the images are divided and sent to different cores to process in the same way. This is a kind of explicit parallelism where the cores are assigned to the same tasks on different portions of the image. This is depicted in Fig. 1.

The different tasks that can be performed using explicit parallelism are contrast stretching, histogram equalization, noise smoothing, filtering, sharpening, etc. The three algorithms considered are histogram equalization, median filtering, and sharpening.

Most of the enhancement algorithms can be improved in performance by explicit parallelism. Different parts of the image are distributed among the cores on the machine and are parallely executed. The dependency of the operations need not be checked as the image is divided into sub-images to do the same task on them.

The method is depicted in Fig. 1. An image is divided into sub-images as per the number of cores on the system. If parallelism is not added into the algorithm, it is executed by the single core and becomes the sequential algorithm. If the number of cores is two, the image can be divided into two different ways like vertical split and horizontal split. In the same way, if the number of cores is four, it can be divided in two different ways as shown in Fig. 1.

Once the algorithm is chosen, and a number of cores of the system are known, it can be run in MATLAB by assigning each of the sub-images to different cores. It can be visualized in pmode where each part is executed by labindex as shown in Fig. 2.

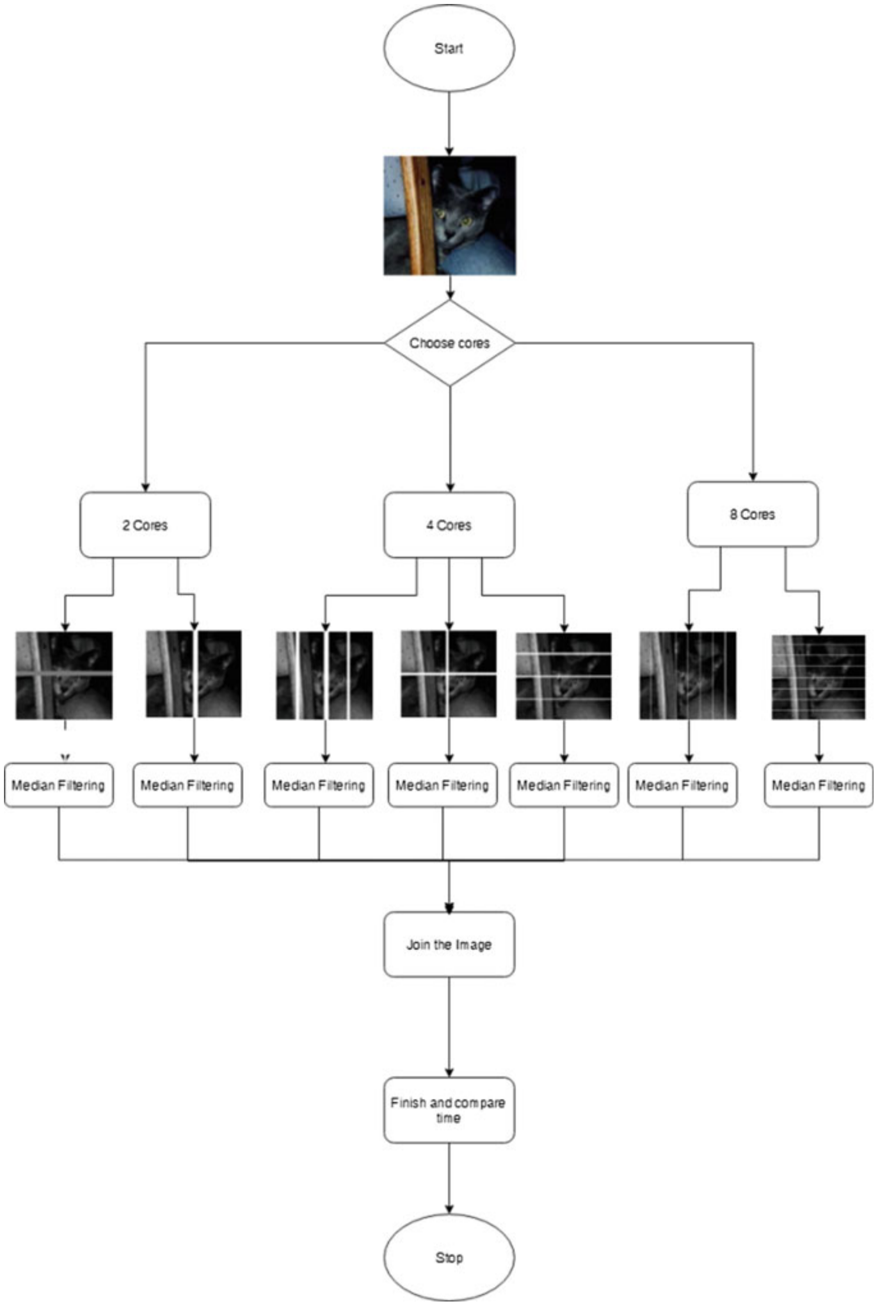


Fig. 1 System model

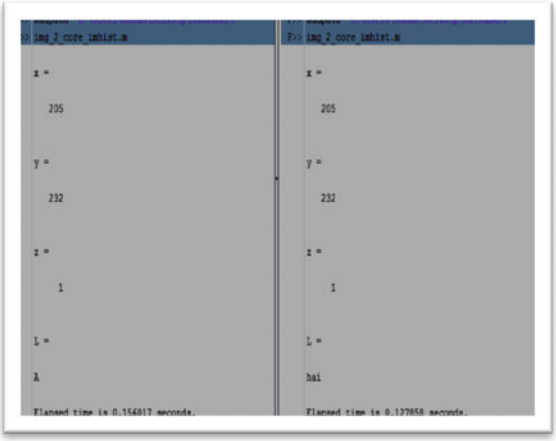


Fig. 2 Sample program running in interactive mode—pmode with two workers/labindex

A labindex is an instance of MATLAB using PCT. When a code is executed in parallel, it is run by the workers/lab indexes than by the main MATLAB. Each labindex is associated with a core of the system. Typically, MATLAB is asked to start one worker for each of the cores in the machine. Each of those workers will be run as a process by the OS, and will end up being run one worker per core in parallel.

Median filters are statistical nonlinear filters that are often described in the spatial domain and are used to remove random noise [10].

Enhancing the image is by finding the median value in a 3 × 3 neighborhood around the corresponding pixel in the input image. It does the filtering of the image in two dimensions. The corresponding pixels on the edges are found with the median values which are substituted with zeros for the missing neighborhood pixel values. As the median is calculated with zero values, the enhanced image may look distorted at the edges.

The histogram equalization and median filtering are the algorithms used for analyzing the explicit parallelism of the proposed model.

The algorithm is executed in a different number of labindex which is chosen according to the number of cores on the machine associated with it. The algorithm is executed on a single labindex with the original image as shown in Fig. 3. This is a sequential execution. The same algorithm is executed in parallel on two parts of the image by two labindex, four parts of the image by four labindex, and eight parts of the image by eight labindex.

The model is tested in the system with following specifications:

CPU—i7-4770 (Quad core) 3.4 GHz,
RAM—4 GB ddr3.

4 Results and Analysis

The low-resolution image considered is an image with the size 255 × 204, as shown in Fig. 3.

Table 1 shows the time taken by the computer to execute the algorithm using a single core (sequential), two cores, four cores, and eight cores for performing image enhancement on this image. The image is partitioned according to the number of cores used for parallelism.

For medium resolution image, the considered image is of dimensions 512 × 512, as shown in Fig. 4. Table 2 contains time taken by the computer to perform the algorithm with this image.

The high-resolution image used for applying this algorithm is of dimensions 1419 × 1001, as shown in Fig. 5. The data computed for this image is given in Table 3.

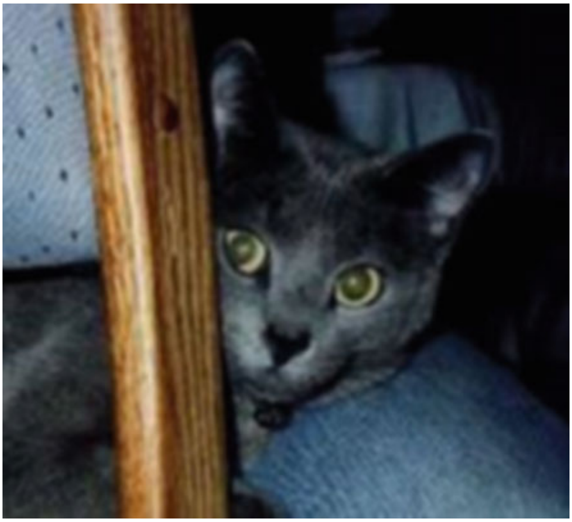


Fig. 3 Low-resolution image

Table 1 Image analysis for low-resolution image

No. of cores	Time taken	Speedup time	Efficiency (%)
1 (Sequential)	0.00404		
2 (Horizontal)	0.00593	0.68213	34.107
2 (Vertical)	0.00417	0.97048	48.524
4 (Horizontal)	0.00900	0.44947	11.237
4 (Vertical)	0.00693	0.58324	14.581
4 (Box)	0.005289	0.76485	19.121
8 (Horizontal)	0.00498	0.81266	10.158
8 (Vertical)	0.00817	0.49498	6.187

Fig. 4 Medium-resolution image

Table 2 Image analysis for medium-resolution image

No. of cores	Time taken	Speedup time	Efficiency (%)
1 (Sequential)	0.00548		
2 (Horizontal)	0.00582	0.94141	47.070
2 (Vertical)	0.01578	0.34723	17.362
4 (Horizontal)	0.00988	0.55450	13.862
4 (Vertical)	0.00964	0.63129	15.782
4 (Box)	0.009642	0.56824	14.206
8 (Horizontal)	0.00657	0.83445	10.431
8 (Vertical)	0.00641	0.85529	10.691

The speedup time from the observed values is given by

$$\text{speedup} = \frac{T_s}{T_p}$$

To find the efficiency, we use the following formula:

$$\text{efficiency} = \frac{T_s}{pT_p},$$

where p is the number of processors used,

T_s and T_p are the time taken to perform the operation in sequential and parallel methods.

Fig. 5 High-resolution image

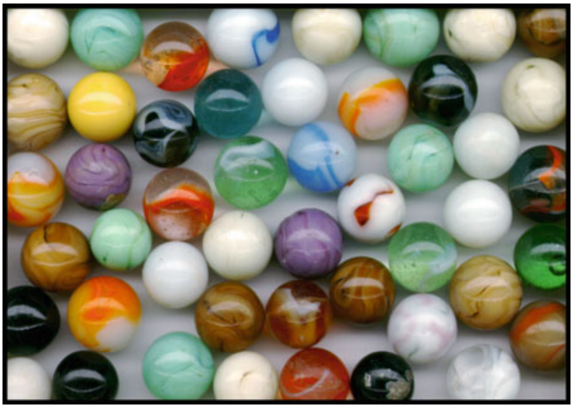


Table 3 Image analysis for high-resolution image

No. of cores	Time taken	Speedup time	Efficiency (%)
1 (Sequential)	0.00548		
2 (Horizontal)	0.01071	0.51167	25.584
2 (Vertical)	0.01227	0.44672	22.336
4 (Horizontal)	0.01564	0.35023	8.756
4 (Vertical)	0.01808	0.30299	7.575
4 (Box)	0.01542	0.35527	8.882
8 (Horizontal)	0.01803	0.30387	3.7998
8 (Vertical)	0.01828	0.29973	3.747

5 Conclusion

Explicit parallelism is tested on image preprocessing algorithms like median filtering in three scenarios with different number of cores. The primary analysis is done on the basis of speedup time and efficiency. It has been observed that the number of partitions required for splitting the job to many cores may degrade the efficiency. The parallelization of the above algorithm has shown that the efficiency is highest when two cores are used. It is observed that the resolution does not show much variation in speedup and efficiency. The same can be tested for various algorithms with more parameters to understand the trade-off between parallelism and the overheads.

Intermediate representation can improve the efficiency of parallelism [11]. The analysis has to be extended with more parameters such as overhead time and fork time to understand the maximum advantage of explicit parallelism in image enhancement algorithms using MATLAB.

References

1. Liu F, Seinstra F, Plaza A (2011) Parallel hyperspectral image processing on distributed multicluster systems. J Appl Remote Sens 5
2. Saxena S, Sharma N, Sharma S (2013) Image processing tasks using parallel computing in multi core architecture and its applications in medical imaging. Int J Adv Res Comput Commun Eng 2(4)
3. Squyres J, Lumsdaine AB, McCandless A, Stevenson R (1996) Parallel and distributed algorithms for high speed image processing
4. Kessler C, Keller J (2007) Models of parallel computing: review and perspectives. PARS-Mitteilungen 24:13–29. ISSN 0177-0454,GI/ITG PARS
5. Alyasseri Z (2014) Survey of parallel computing with MATLAB. Cornell University Library, arXiv:1407.6878, Article
6. Kaur P (2015) Implementation of image processing algorithms on the parallel platform using MATLAB. Int J Comput Sci Eng Technol (IJCSET)
7. Saxena S, Sharma S, Sharma N (2016) Parallel image processing techniques, benefits and limitations. Res J Appl Sci Eng Technol 12(2):223–238
8. Ranjith R, Shanmughasundaram R (2015) Simulation of safety critical applications for automotive using multicore scheduling. In: International conference on control, instrumentation, communication and computational technologies (ICCICCT)
9. Chatarasi P, Shirako J, Sarkar V (2015) Polyhedral optimization of explicitly parallel programs. In: International conference on parallel architecture and compilation (PACT), pp 213–226
10. Saxena C, Kourav D (2014) Noises and image denoising techniques: a brief survey. Int J Emerg Technol Adv Eng 4(3)
11. Belwal M, Sudarshan TSB (2015) Intermediate representation for heterogeneous multi-core: a survey. In: International conference on VLSI systems, architecture, technology and applications (VLSI-SATA)

A Comprehensive Study on Character Segmentation

Sourabh Sagar and Sunanda Dixit

Abstract In identifying the characters from a given image, character segmentation plays an important role. In a given line of text, first, we have to segment the words. Then, in each word there will be a character-by-character segmentation. There have been some rapid developments in this area. Many algorithms have been implemented to increase the accuracy range and decrease the word error rate. This paper aims to provide a review of some of the developments that have happened in this domain.

1 Introduction

Lot of research work has been going on, for some time now. Results of this research have increased the accuracy range from research to further. But still there is no method which produces 100% perfect segmentation. Because the cursive handwriting does not have any correct format. People have their own style of writing. So, it will be a very difficult task to compare the input image with its database. This paper provides a review on some of the methods which are used for character segmentation.

2 Literature Survey

Manjunath and Sharath [1] provide an application that can run on any android device because they are portable. This paper uses Kohonen's algorithm. It will extract the text from the image and finds the meaning on the Internet. Zhu et al. [2] use feature coordinates as unary features. For binary features, this method uses the differences

S. Sagar (✉) · S. Dixit
Department of Information Science and Engineering,
Dayananda Sagar College of Engineering, Bangalore, India
e-mail: sourabhacademic@gmail.com

S. Dixit
e-mail: sunanda.bms@gmail.com

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_141

in coordinates between the neighboring feature points. Kozielski et al. [3] proposed the conversion of digit recognition to handwritten recognition. Independent normalization procedure is applied for the image slices. Additional algorithm used here is a novel moment-based algorithm. Wu et al. [4] maintained a codebook containing SIFT descriptors (SDs). Here, shift descriptors signature (SDS) and scale and orientation histogram (SOH) are used to identify the writing. Six public datasets are used for this experiment. Tan et al. [5] proposed a method where strokes are generated for each line by factorizing to get the similarity matrix. To get labels for strokes, nonlinear clustering methods are applied to the similarity matrix.

Ghosh and Valveny [6] use a simplified version of an attribute model to get an integral image representation of the input document. It increases the efficiency. Performance will be increased by the introduction of re-ranking step. Naveena and Manjunath Aradhya [7] proposed a method where mixture of Gaussians is learnt by applying the expectation maximization (EM) algorithm. For segmenting the characters, branch points are used as reference points. Given results are satisfactory. Haji et al. [8] proposed an algorithm which can be complemented with any existing segmentation algorithm. More accuracy in segmentation rate is achieved. Parui et al. [9] proposed an algorithm. Script and particular intelligent segmentation strategy are independent of this algorithm. Based on different scripts, different segmentation algorithms are explored. HMM algorithm is evaluated. Chen et al. [10] mainly focus on concatenating the small segments in order to overcome the over segmentation.

Lawgali et al. [11] provide huge database of 6600 shapes of 50 writers. The name of database is handwritten Arabic character database (HACDB) which will be useful for the recognition of characters after segmentation. Sharma et al. [12] provided the method which increases the recognition accuracy in minimum amount of time. Using Euler number, the search space will be reduced. Saba et al. [13] focus on unconstrained handwritten words. It figures out the characters and compares it with the database. Artificial neural network is employed. Marti et al. [14] provide a database for English language named Lancaster–Oslo/Bergen (LOB). It will be helpful for the recognition of characters after segmentation.

Prasanna et al. [15] approach help the blind or visually impaired person. This approach will identify the characters in the taken image and gives output in the form of voice. So, a blind person can get the information from that output through hearing. Sharma and Sharma [16] proposed that identifying characters in an image is an easy task for humans but not for machines. Humans use their previous knowledge and compare the characters but the machine cannot do. Sandhya and Krishanan [17] have provided algorithm for reconstruction of the broken characters and recognition. In historical documents, the characters will be broken and blurred. In those cases, it will be used. Pardeshi et al. [18] used the wavelet-like discrete cosine transform (WDCT) to separate the handwritten words and printed words. Banumathi and Jagadeesh Chandra [19] proposed projection profile technique for character segmentation comprising preprocessing, binarization, and many other processes. Patil and Hanni [20] proposed the optical character recognition (OCR) which segments the whole document into text lines, then words, and then characters. Those characters are used for recognition.

The input for character recognition is the output of line segmentation, while doing a survey the following techniques are observed on text line segmentation.

Dixit et al. [21] proposed a new cost function which considers the line spacing and skew of each line. For baseline correction, a new algorithm is used which improves the efficiency. Dixit and Suresh [22, 23] proposed a method where adaptive histogram equalization and sliding window techniques play a very important role. This is compared with conventional text line segmentation technique which leads to an improvement. Dixit et al. [24] proposed a method for line segmentation, where based on hough lines, text lines are identified. Noise will be removed after the segmentation result. Dixit and Narayan [25] discuss a novel technique for line segmentation. Preprocessing like skew correction is done.

Survey Table

Year	Title	Methodologies used	Pros	Cons	Language
2016	Character recognition using image processing [16]	Optical character recognition (OCR)	Provides an upper edge	Accuracy depends on material	English
2016	Broken Kannada character recognition [17]	End-point algorithm	Achieves 98.9% accuracy		Kannada
2016	Line and word segmentation of Kannada handwritten text documents using projection profile technique [19]	Projection profile technique	90% successful segmentation	Word segmentation is not satisfactory	Kannada
2016	Handwritten and machine printed text separation from Kannada document images [18]	Wavelet-like discrete cosine transform (WDCT), k-NN classifier	Achieves 99.50% accuracy		Kannada
2016	Handwritten Kannada document image processing using optical character recognition [20]	Optical character recognition (OCR)	Measurable improvement	Skew detection and noise removal techniques can still be improved	Kannada
2015	Script independent online handwriting recognition [9]	Hidden Markov model (HMM)	Three segmentation strategies		English, Arabic, Bangla, Bengali
2015	Query by string word spotting based on character bigram indexing [6]	Pyramidal histogram of characters (PHOC)	Segmentation free	Retrieval method needs improvement	English

(continued)

(continued)

Year	Title	Methodologies used	Pros	Cons	Language
2014	Offline text-independent writer identification based on scale-invariant feature transform [4]	Scale-invariant feature transform (SIFT)	Writer identification	Segmentation is not easy	English, Chinese
2014	Segmentation of Kannada handwritten text line through computation of variance [25]	Computation of variance			
2014	Kannada text line extraction based on energy minimization and skew correction [21]	Skew correction, background cleaning	Segment the lines of text		Kannada
2014	Text line segmentation of handwritten documents in Hindi and English [24]	Hough line	Non-skewed also considered		Hindi, English
2014	Sliding window technique for handwritten text line segmentation [23]	Sliding window, OCR, adaptive histogram equalization		Not preferable for images with complex characteristics	Kannada
2013	South Indian Tamil language handwritten document text line segmentation technique with aid of sliding window and skewing operations [22]	Sliding window, adaptive histogram equalization	Better method than the conventional		Tamil
2013	Implementing Kannada optical character recognition on the Android operating system for Kannada signboards [1]	Kannada OCR, Kohonen's algorithm, Hilditch algorithm	Portable, easy to use, recognize, translation	Less accurate, do not support for more languages	Kannada
2013	Online handwritten cursive word recognition using segmentation-free MRF in combination with P2DBMN-MQDF [2]	Markov random field (MRF), Pseudo 2D bi-moment normalization (P2DBMN), Modified quadratic discriminant function (MQDF)	Restricts the search space	More time is needed	Chinese, Japanese

(continued)

(continued)

Year	Title	Methodologies used	Pros	Cons	Language
2013	Keyword spotting in unconstrained handwritten Chinese documents using contextual word model [10]	Contextual word model	Higher recall rate		Chinese
2013	Handwritten Arabic character database (HACDB) for automatic character recognition [11]	Building their own database	Large database		Arabic
2012	Moment-based image normalization for handwritten text recognition [3]	Normalization process, a novel moment-based algorithm	Achieved word error rate to 13.4% from 16.7%	Language model's perplexity is 258.7	French, English
2012	A new handwritten character segmentation method based on nonlinear clustering [5]	Spectral clustering based on normalized cut (Ncut), Kernel clustering based on conscience online learning (COLL)	Segment nonlinearly separable characters	Took few databases	English, Chinese
2012	Handwritten character segmentation for Kannada scripts [7]	Expectation maximization (EM) algorithm, cluster mean points	Satisfactory result	Not good for overlapping components	Kannada
2012	Statistical hypothesis testing for handwritten word segmentation algorithms [8]	Hidden Markov model (HMM)	Can be used with any segmenta-tion algorithm	Segmentation is not perfect	English
2012	An improved zone-based hybrid feature extraction model for handwritten alphabets recognition using Euler number [12]	Euler number	Increases speed and accuracy	Not a complete mechanism	English
2011	Cursive script segmentation with neural confidence [13]	ANN, CEDAR database	Accuracy is increased		

(continued)

(continued)

Year	Title	Methodologies used	Pros	Cons	Language
2011	Kannada text extraction from images and videos for vision-impaired persons [15]	Text extraction, text recognition, speech synthesis	Blind people can get the information		Kannada
1998	A full English sentence database for offline handwriting recognition [14]	Database	A basis for character recognition		English

3 Conclusion

There has been a lot of advancement in character segmentation domain. Yet, no perfect segmentation technique is found. But, there are chances of getting perfect segmentation which shows 100% accuracy. Through this survey, we have come to realize that there are several segmentation techniques existing. Applying these different techniques, we can try to achieve good accuracy level in this domain.

References

1. Manjunath AE, Sharath B (2013) Implementing Kannada optical character recognition on the android operating system for Kannada sign boards. IJARCCE (Int J Adv Res Comput Commun Eng) 2(1)
2. Zhu B, Shivram A, Setlur S, Govindaraju V, Nakagawa M (2013) Online handwritten cursive word recognition using segmentation-free MRF in combination with P2DBMN-MQDF. IEEE, pp 349–353
3. Kozielski M, Forster J, Ney H (2012) Moment-based image normalization for handwritten text recognition. IEEE, pp 256–261
4. Wu X, Tang Y, Bu W (2014) Offline text-independent writer identification based on scale invariant feature transform. IEEE 9(3)
5. Tan J, Lai JH, Wang CD, Wang WX, Zuo XX (2012) A new handwritten character segmentation method based on nonlinear clustering. Elsevier, pp 213–219
6. Ghosh SK, Valveny E (2015) Query by string word spotting based on character Bi-Gram indexing. IEEE, pp 881–885
7. Naveena C, Manjunath Aradhya VN (2012) Handwritten character segmentation for Kannada scripts. IEEE, pp 144–149
8. Haji M, Sahoo KA, Bui TD, Suen CY, Ponson D (2012) Statistical hypothesis testing for handwritten word segmentation algorithms. IEEE, pp 114–119
9. Samanta O, Roy A, Bhattacharya U, Parui SK (2015) Script independent online handwriting recognition. IEEE, pp 1251–1255
10. Huang L, Yin F, Chen Q-H, Liu C-L (2013) Keyword spotting in unconstrained handwritten Chinese documents using contextual word model. Image Vis Comput 31:958–968
11. Lawgali A, Angelova M, Bouridane A (2013) HACDB: handwritten Arabic characters database for automatic character recognition. IEEE

12. Sharma OP, Ghose MK, Shah KB (2012) An improved zone based hybrid feature extraction model for handwritten alphabets recognition using euler number. IJSCE (Int J Soft Comput Eng) 2(2). ISSN: 2231-2307
13. Saba T, Rehman A, Sulong G (2011) Cursive script segmentation with neural confidence. IJIC (Int J Innov Comput) 7(8)
14. Marti UV, Bunke H (1998) A full English sentence database for off-line handwriting recognition. IEEE
15. Prasanna K, Ramakhanth Kumar P, Thungamani M, Koli M (2011) Kannada text extraction from images and videos for vision impaired persons. IJAET (Int J Adv Eng Technol). ISSN: 2231-1963
16. Sharma S, Sharma R (2016) Character recognition using image processing. IJAETMAS (Int J Adv Eng Technol Manage Appl Sci) 03(09):115–122. ISSN:2349-3224
17. Sandhya N, Krishanan R (2016) Broken Kannada character recognition—a neural network based approach. In: ICEEOT (International conference on electrical, electronics and optimization techniques)
18. Pardeshi R, Hangarge M, Doddamani S, Santosh KC (2016) Handwritten and machine printed text separation from Kannada document images. IEEE
19. Banumathi KL, Jagadeesh Chandra AP (2016) Line and word segmentation of Kannada handwritten text documents using projection profile technique. In: ICEECCOT (International conference on electrical, electronics, communication, computer and optimization techniques)
20. Patil MM, Hanni AR (2016) Handwritten Kannada document image processing using optical character recognition. IOSR-JCE (IOSR J Comput Eng) 18(4), Ver. VI:39–47. e-ISSN:2278-0661, p-ISSN:2278-8727
21. Dixit S, Hosahalli Narayan S, Belur M (2014) Kannada text line extraction based on energy minimization and skew correction. In: IEEE IACC (International advance computing conference)
22. Dixit S, Suresh HN (2013) South Indian tamil language handwritten document text line segmentation technique with aid of sliding window and skewing operations. JATiT (J Theoret Appl Inf Technol) 58(2). ISSN: 1992-8645, E-ISSN: 1817-3195
23. Dixit S, Suresh HN (2014) Sliding window technique for handwritten text line segmentation. IJRCEE (Int J Res Comput Eng Electron) 3(04). ISSN: 2319-376X
24. Dixit S, Sneha NU, Suresh HN (2014) Text line segmentation of handwritten documents in Hindi and English. IJRITCC (Int J Recent Innov Trends Comput Commun) 02(04). ISSN: 2321-8169
25. Dixit S, Narayan SH (2014) Segmentation of Kannada handwritten text line through computation of variance. IJCSIS (Int J Comput Sci Inf Sec) 12(02). ISSN: 1947-5500

EZW, SPIHT and WDR Methods for CT Scan and X-ray Images Compression Applications

S. Saradha Rani, G. Sasibhushana Rao and B. Prabhakara Rao

Abstract Scanning rate of medical image tools has been significantly improved owing to the arrival of CT, MRI and PET. For medical imagery, storing in less area and not losing its details are vital. So, an efficient technique is necessary for storing in a cost-effective way. In this paper, wavelet is employed to perform decomposition, and image is compressed using Embedded Zero-Tree Wavelet (EZW), Set Partitioning in Hierarchical Trees (SPIHT) and Wavelet Difference Reduction (WDR) algorithms. These algorithms are applied to compress X-ray and CT images, and compared using performance metrics. From results, it is seen that compression ratio is better in WDR for all the wavelets than SPHIT and EZW. High compression ratio, 82.47, is obtained with Haar and WDR combination for CT scan, whereas this is 32.89 for Biorthogonal and WDR combination for X-ray. The main objective of this paper is to find the optimal combination of wavelets and image compression techniques.

1 Introduction

These days, the image compression algorithms for decreasing the dimensions play a significant task in an image processing predominantly in communications, visualizations, classifications of image and storage of high-dimensional information. The majority of the mentioned applications are impracticable with no compression [1].

S. Saradha Rani (✉)
Department of Electronics and Communication Engineering, GITAM, Visakhapatnam 530045, Andhra Pradesh, India
e-mail: ssaradarani@gmail.com

G. Sasibhushana Rao
Department of Electronics and Communication Engineering, AU College of Engineering, Andhra University, Visakhapatnam 530003, India
e-mail: sasigps@gmail.com

B. Prabhakara Rao
Department of Electronics and Communication Engineering, JNTUK, Kakinada 533003, Andhra Pradesh, India
e-mail: drbprjntuk@gmail.com

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_142

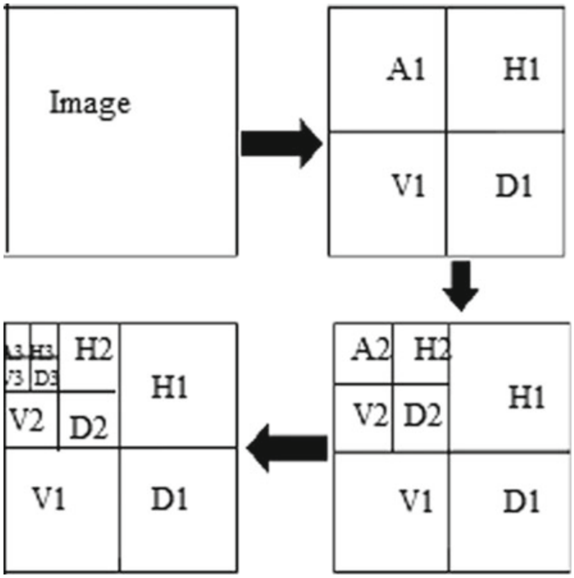


Fig. 1 Process of three-level wavelet decomposition

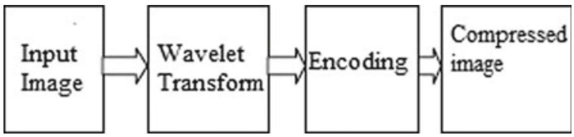


Fig. 2 Block diagram of image compression

Compression is a procedure of lessening the quantity of required bits to signify the data or image, without diminishing the nature of the image. By compressing the data, it is possible to accumulate storage space capability; reducing expenditure for network bandwidth and storage equipment can boost up the transfer of file. Compression is a means of describing the data in a conservative form fairly than its original form. The process of wavelet decomposition involves low-pass and high-pass filtering. The wavelet transform partitions the image into numerous sub-bands, namely, HH, HL, LL and LH; among these sub-bands, merely LL sub-band is further decomposed, as it has low-frequency component and noise compared to the remaining sub-band levels [2].

The process of wavelet decomposition is shown in Fig. 1.

The compression is performed in two levels. First, the wavelet transform is implemented to decompose the image; second, encoding is applied for compressing the image and is illustrated with block diagram shown in Fig. 2 [3]. In this paper, Haar, Daubechies and Biorthogonal wavelets have been used for decomposition, and EZW, SPIHT and WDR algorithms have been applied for encoding.

2 Embedded Zero-Tree Wavelet (EZW)

In wavelet-based compression, EZW method was one in all the initial and dominant algorithms. Further algorithms were shaped relying on the elementary concept of EZW [4]. The heart of EZW compression is the development of self-resemblance across the distinct image wavelet transform scales [5]. Alternatively, in wavelet decomposed image, EZW approximates the high occurrence coefficients. For the reason that the wavelet coefficients have data concerning each frequency and spatial content of a picture, removal of a high-frequency coefficient results in a minor degradation of image in a specific position of the restored image instead of that across complete picture. In this, the threshold is employed to compute a significance map of significant and insignificant coefficients. The significance map can be represented with zero trees efficiently [5]. The following are the main steps in EZW:

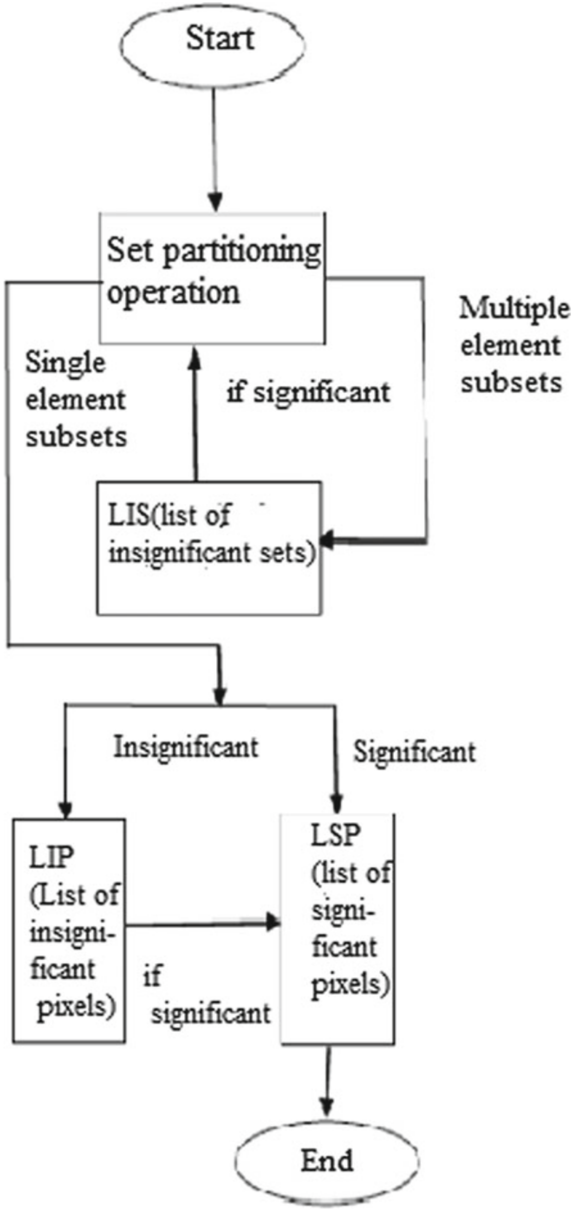
(a) Initializing threshold: Threshold T is initialized to the lowest power of 2 in which 'T' exceeds $|C_{\max(k,1)}|/2$, where $C_{k,1}$ is wavelet coefficients.
(b) Coding of significance map: when $|C_{k,1}| > T$, output a symbol by scanning the coefficients in a predefined manner. When this symbol inputs the decoder, it assigns $C_{k,1} = \pm 1.5T$.
(c) Refinement: in this step, every significant coefficient is refined by transmitting the next binary bit. Decoder augments the present value by $\pm 0.25T$, as soon as it receives [5].
(d) Put $T = T/2$, as well as repeat from (b) if further iterations are desired.

3 Set Partitioning in Hierarchical Trees (SPIHT)

The SPIHT encoding algorithm is developed edition of EZW. For different types of images, high values of PSNR can be achieved using SPIHT technique. It offers an improved assessment standard for successive techniques. Hierarchical trees pertain to quadtrees, described in EZW [6]. Set partitioning means the manner how quadtrees separate or split, and the wavelet transforms at a certain threshold. The coding technique uses three lists, known as List of Insignificant Sets (LIS), List of Significant Pixels (LSP) and List of Insignificant Pixels (LIP) [6]. The subsequent process illustrates the encoding using SPIHT:

(a) Initialization: set Output $n = [\log_2 {}^{C\max(k,1)}]$, LSP = { }, LIP = {(k, 1)} and LIS = {$D(k, 1)$} [5].
(b) Sorting pass: considering every access in LIP, if it is significant to produce an output as '1' by removing it from LIP and adding to LSP, otherwise output as '0'. Output the significance, i.e. output its sign if significant, for every access in the LIS. Carry out set partitioning, relying upon whether it is the $D(k, 1)$ or the $L(k, 1)$ set. And update the LSP, LIP and LIS accordingly [7].
(c) Refinement pass: on considering every access in LSP, apart from those that are brought all through the sorting pass with similar 'n', encode the nth MSB [8].

Fig. 3 SPIHT algorithm flow chart



(d) In present pass, *n* is decremented by ‘1’ and the steps (b) and (c) are repetitive till ‘*n*’ reduces to zero.

The SPIHT algorithm is illustrated in flow chart shown in Fig. 3.

4 Wavelet Difference Reduction (WDR)

The imperfection in SPIHT is that it solely finds the location pertaining to significant coefficients, making hard to carry out the operations like Region of Interest (ROI) on compressed data, which rely on significant transform values precise location. That means a portion that requires increased resolution is to be selected in compressed image. WDR technique is completely easy procedure. First, an image is transformed using wavelet. Second, for the wavelet coefficients, bit-plane-based WDR encoding that is based on bit-plane coding is applied. The main five steps that constitute WDR are as follows:

(a) Initialization: select initial value of threshold ‘T_0’ so that ‘T_0’ is superior to all transformed values and at least one transform value has a level of $T_0/2$ [9].
(b) Threshold updation: assign $T_n = T_{n-1}/2$.
(c) Significance pass: for the deemed significant values, i.e. transform values that are equal or more than the threshold value, index values will be encoded by difference reduction method [10]. The difference reduction technique basically encodes the range of steps required to shift from previous significant value index to the present significant value index [10]. Significance pass output includes bit sequence produced by difference reduction and significant values signs, describing the accurate positions of significant values [10].
(d) Refinement pass: similar to the process in SPIHT method, here it generates refined bits by using standard bit- plane quantization procedure [10]. These refined bits are the improved approximation of the exact transform bits.
(e) Perform (b)–(d).

5 Performance Measures

5.1 Compression Ratio, CR

It is the quantitative relation between amount of bits required for representation of true picture to the amount of bits in compressed image [9]. The expression is

$$CR = \frac{n_1}{n_2} \tag{1}$$

where

n_1 amount of bits in true l image and
n_2 amount of bits in compressed image.

5.2 Mean Squared Error, MSE

MSE is the error metric accustomed to measure up to the excellence of different image compression methods [9].

$$\text{MSE} = \frac{1}{mn} \sum_{k=0}^{m-1} \sum_{l=0}^{n-1} (O(k,l) - C(k,l)) \quad (2)$$

where $m\ X\ n$ is dimension of the image, $O(k,\ l)$ is true image and $C(k,\ l)$ is the approximated or compressed image.

5.3 Peak Signal-to-Noise Ratio, PSNR

The other error metric is PSNR that is used to compute up to the subjective faithfulness of the uncompressed image [9]. PSNR is the measure of quality of the compressed image.

$$\text{PSNR} = 10 \log_{10}\left(\frac{m\text{X}n}{\text{MSE}}\right) \quad (3)$$

6 Experimental Results

The results have been analysed on test image and two medical images. The wavelets Haar, Daubechies and Biorthogonal have been used to analyse the methods EZW, SPIHT and WDR, and various performance measures PSNR, MSE, CR and BPP are calculated for each of the methods for the following images.

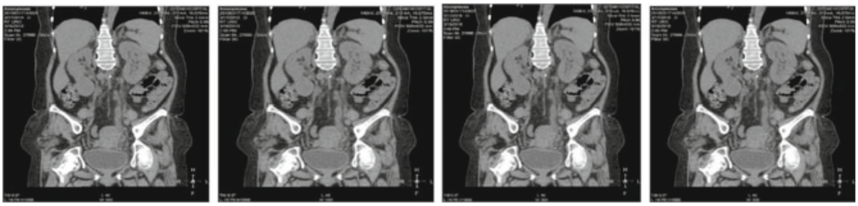
Figure 4 illustrates the original image and compressed images of test image, and Figs. 5 and 6 depict the compressed images of CT scan of lower abdomen and X-ray of shoulder.

(a) original image (b) EZW Compression (c) SPIHT Compression (d) WDR Compression

Fig. 4 Compression of African sculpture test image using various image compression methods

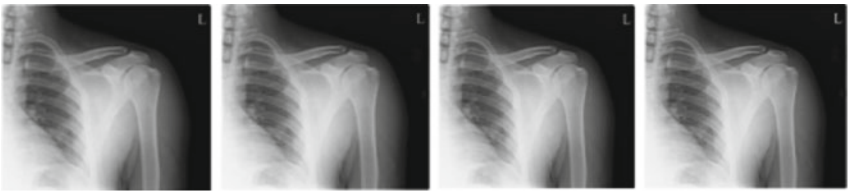
Table 1 Performance analysis of various algorithms for different wavelet and medical image configurations

Wavelet	Algorithm	CT scan of lower abdomen				X-ray of shoulder			
		C.R	PSNR	MSE	BPP	C.R	PSNR	MSE	BPP
HAAR	EZW	69.95	46.0833	0.0447	16.7876	26.21	54.66	0.2222	2.0972
	SPIHT	40.08	45.15	1.987	9.6201	10.75	45.78	1.718	0.8599
	WDR	82.47	46.08	1.602	19.7919	31.6	49.77	0.6858	2.5278
DAUBECHIES	EZW	67.79	61.46	0.0464	16.2706	19.92	54.5	0.2306	1.5934
	SPIHT	37.1	44.98	2.067	8.9034	7.88	46.52	1.448	0.6306
	WDR	79.1	46	1.634	18.1843	23.23	50.02	0.6478	1.8585
BIORTHOGONAL	EZW	66.1	60.45	0.0586	15.8832	27.44	54.51	0.2302	2.1948
	SPIHT	35.22	45.54	1.818	8.452	11.39	45.42	1.866	0.9112
	WDR	77.22	45.77	1.357	18.5356	32.89	48.22	0.979	2.631



(a) original image (b) EZW Compression (c) SPIHT Compression (d) WDR Compression

Fig. 5 Compression of CT scanned lower abdomen image using various image compression methods



(a) original image (b) EZW Compression (c) SPIHT Compression (d) WDR Compression

Fig. 6 Compression of X-ray of shoulder image using various image compression methods

Table 1 shows the performance measures of EZW, SPIHT and WDR algorithms. From the table, it is seen that the compression ratio is better for WDR algorithm for all the wavelets. For CT scanned lower abdomen image, the best measures are C.R = 82.47 and PSNR = 46.08 for Haar wavelet and for X-Ray of shoulder, C.R = 32.89 for Biorthogonal wavelet and PSNR = 49.77 obtained for Haar wavelet.

7 Conclusion

In this paper, an assessment of the discussed algorithms is carried out on CT and X-ray images. Here, the image is decomposed using Haar wavelet. After transforming, it is encoded using different algorithms. The same procedure is repeated using Daubechies and Biorthogonal wavelets. The capability of the algorithms to compress an image is analysed in terms of CR, MSE, BPP and PSNR. Quality of image is calculated by means of Mean Squared Error (MSE) and Peak Signal-to-Noise Ratio (PSNR), while compression measure is analysed via compression ratio and bits per pixel.

Acknowledgements The author would like to thank GIMSR, Visakhapatnam, for providing lower abdomen CT scan and shoulder X-ray images to carry out the research work.

References

1. Kumari L, Pandian R, Aran Glenn J (2016) Analysis of multi scale features of compressed medical images. In: International conference on electrical, electronics, and optimization techniques (ICEEOT)
2. Negahban Farnoosh, Shafieian Mohammad Ali (2013) Various novel wavelet based image compression algorithms using a neural network as a predictor. Basic Appl Sci Res 3(6):280–287
3. Gonzalez RC, Woods RE (2009) Digital image processing. Pearson Education, 2nd edn
4. Shapiro JM (1997) Embedded image coding using zero trees of wavelet coefficients, pp 108–121
5. Raja SP, Suraliandi A (2010) Performance evaluation on EZW & WDR image compression techniques. In: International conference on communication control and computing technologies
6. Raja SP (2009) Wavelet based image compression: a comparative study. In: 2009 international conference on advances in computing control and telecommunication technologies
7. Said A, Pearlman WA (1993) Image compression using the spatial-orientation tree. In: IEEE international symposium on circuits and systems, Chicago
8. Said A, Pearlman WA (1996) A new, fast, and efficient image codec based on set partitioning in hierarchical trees. IEEE Trans Circ Syst Video Technol 6(3)
9. Vaish A, Kumar M (2015) WDR coding based image compression technique using PCA. In: 2015 international conference on signal processing and communication
10. Singh P, Singh P (2012) Implementation of SPIHT and WDR algorithms for natural and artificial images using wavelets. In: Fourth international conference on computational intelligence and communication networks

Human Identification Based on Ear Image Contour and Its Properties

P. Ramesh Kumar, K. L. Sailaja and Shaik Mehatab Begum

Abstract Identity management is the process of authenticating individuals by means of security objects (traits) to confirm whether the subject is permitted to access any secured property. Ear biometrics is one of the best solutions to access any secured property, which may be private/public. In the current security surveillance, the subject is identified passively without the knowledge. Ear recognition is a better passive system where the human ear is captured to verify whether he is authorized or not. This system can possibly suit for crowd management like bus stations, railway stations, temples, cinema theatres, etc. An ear biometric system based on 2D ear image contours and its properties was proposed. In this article, three types of databases are taken as input, i.e. IIT Delhi Database, AMI Database and VR Students Sample Database, and enrolment and verification process is done with these databases based on the contour features and its properties—bounding rectangle, aspect ratio, extent, equivalent diameter, contour area, contour perimeter, checking convexity, convex hull and solidity. This approach takes less time to execute, and the obtained FAR and FRR performance parameter values are nominal when compared to other traditional mechanisms.

1 Introduction

Biometrics is a method of utilizing the physical parts of a person as a stable durable secret key. Like the fingerprints which are unique for each individual, the face, eyes, ears, hands and voice are also additionally one of a kind. Innovation has progressed to the point where PC frameworks can trace and identify the patterns, hand shapes, ear contours and other physical attributes. Utilizing biometrics gadgets are empowered

P. Ramesh Kumar (✉) · K. L. Sailaja · S. Mehatab Begum
Department of Computer Science & Engineering, VR Siddhartha Engineering College, Vijayawada, India
e-mail: send2rameshkumar@gmail.com

K. L. Sailaja
e-mail: sailaja0905@gmail.com

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_143

with the capacity to check personality instantly and deny access to every other person. Utilizing biometrics for distinguishing and confirming people offers extraordinary points of interest over general customary techniques like smart cards, magnetic stripe cards, etc.

Passwords in the form of text can be overlooked or inadvertently seen by somebody lacks secrecy which makes an issue for clients. Since the biometric confirmation depends on the recognizable proof of a physical part of a person, the individual himself can act as a password in a biometric system. While choosing a biometric type, it is important to give due thought to acceptability. In this sense, the ear is a standout amongst the most appropriate contender to be utilized for biometrics. It does not change with emotions, states of mind, sadness, fear or cosmetic changes. The ear is effectively caught from a distance through a camera, regardless of whether the subject is not completely agreeable.

This makes ear-based identification [1–3] more suitable for passive intelligent security system and for crime investigation. Also, it might be understood that most of the existing biometrics system requires a large hardware set-up in a restricted background. Most of the biometrics systems require automatic recognition algorithm to identify the people based on physical, behavioural or chemical character. The researchers working on the stable biometric attributes and algorithms [3–9] have been explored broadly in recent times. Our assessment considers the deployment of ear as a biometric for human identification. In addition to these, other criteria are added such as feature extraction of ear image and its contours properties are calculated. In this paper, nine features: bounding rectangle, aspect ratio, extend, equivalent diameter, contour area, contour perimeter, checking convexity, convex hull and solidity are extracted. Human identification is done with the support of these contour features.

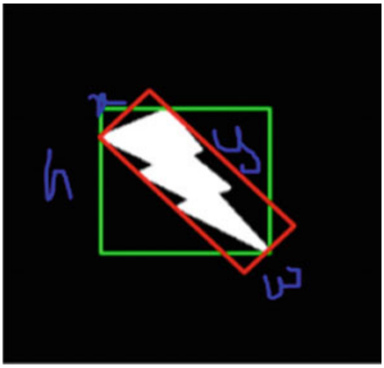
2 Feature Extraction

The proposed approach recognizes the human identity based on ear image contours and reduces the False Rejection Rate (FRR) and False Acceptance Rate (FAR). Here, only ear images are considered, and these images are characterized by the contour features [bounding rectangle, aspect ratio, equivalent diameter, contour area, contour perimeter, checking convexity, convex hull and solidity]. Feature values are calculated, and based on these values, the individuals are easily identified and verified.

2.1 Feature Template

The feature extraction of ear image contours is defined based on some properties, and the following features are extracted [10]:

Fig. 1 Bounding rectangle



1. Bounding rectangle,
2. Aspect ratio,
3. Extent,
4. Equivalent diameter,
5. Contour area,
6. Contour perimeter,
7. Checking convexity,
8. Convex hull and
9. Solidity.

2.1.1 Bounding Rectangle

The Minimum Bounding Rectangle (MBR) is an expression of the maximum extents of a 2-D object like point, line, polygon, etc., or set of objects within its 2-D coordinate system.

The bounding rectangles are classified as follows:

(i) Straight bounding rectangle which never considers the rotation of the thing.
(ii) Rotated rectangle which considers the rotation of the thing (Fig. 1).

2.1.2 Aspect Ratio

It is the ratio between the width and height of an image.

$$\text{Aspect Ratio} = \text{Width}/\text{Height}$$

Width: rectangle width, and Height: rectangle height.

2.1.3 Extent

The area covered by something is called Extent.

$$\text{Extent} = \text{Object Area}/\text{Bounding Rectangle Area}$$

Bounding Rectangle Area:	Area of the rectangle (W * h), and
Object Area:	The area covered by the ear image.

2.1.4 Equivalent Diameter

Here, we can find the diameter of irregular-shaped objects by applying some pressure.

$$\text{Equivalent Diameter} = \sqrt{4} * \text{Contour Area}/\Pi$$

Contour Area: Bounding area of the ear image.

Contour

It is an outline representing or bounding the shape or form of something. It moulds into a specific shape, especially one designed to fit into something else.

2.1.5 Contour Area

An outline representing or bounding the shape or form of something is called contour, and the regional area is called contour area.

$$\text{Area} = \text{contourArea()}$$

2.1.6 Contour Perimeter

The contour perimeter or arc length is the continuous line forming the boundary of a closed figure.

$$\text{Perimeter} = \text{arcLength (contour)}$$

2.1.7 Checking Convexity

This property is to check whether the curve is convex or not.

$$\text{Checking Convexity} = \text{is Contour Convex (contour)}$$

2.1.8 Convex Hull

convexHull() is used to check the curve for convexity defects and correct it.

2.1.9 Solidity

The quality or state of being firm or strong in structure. Example: wrapping a rubber band around the region.

$$\text{Solidity} = \text{Contour Area}/\text{Convex Hull Area}$$

3 Methodology

The architecture of this methodology consists of five steps (Fig. 2). Initially, raw image is taken from image acquisition device. The second step is preprocessing, where the ear image is preprocessed manually or with the help of cropping tools such as Matlab, snipping tool, etc. The third stage is image binarization. In this stage, binarized image will be obtained. In the fourth and final stage, the contour features are extracted, and the individual can be easily identified and verified by comparing with the created feature database.

1. Input image acquisition,
2. Preprocessing for ear image,
3. Image binarization,
4. Contour feature extraction and
5. Comparison.

Input Image Acquisition

The input images are taken from three databases:

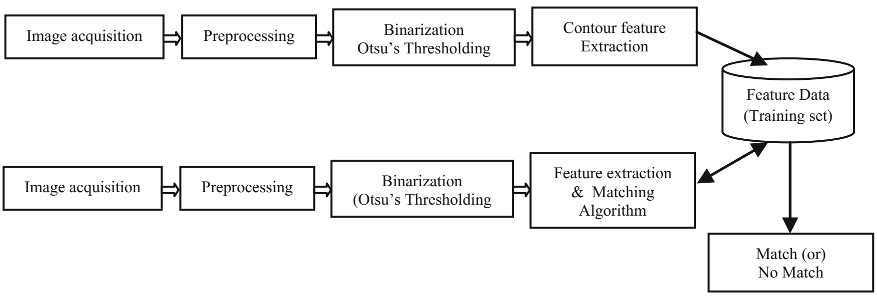


Fig. 2 Enrolment and verification process

1. IIT Delhi Database.
2. AMI Database.
3. VR Students Sample Database.

Preprocessing for Ear Image

In the preprocessing stage, the ear image which is extracted from the side view of the face gets preprocessed by removing shadows, hair, ornaments, etc., around the ear image. The ear shape is extracted by applying a series of grayscale morphological operations.

Image Binarization

The preprocessed ear image is used to develop a binarized mask which covers the required region for feature extraction using Otsu's threshold. The obtained image produces masks with different dimensions. The thresholding limit is adjusted to make the mask area less than the predefined limit.

Feature Extraction

The contour of an interested region in an image helps the image analyst to extract the shape of the object. The contours are good features for object detection and recognition. The best suitable image for contour processing is binary image, and any edge detection or threshold method can be easily applied on it to find the contours. The nine different contour features and their properties are extracted here.

Comparison

In this step, first perform image acquisition and in the preprocessing stage raw images are cropped and preprocessed images are used in the feature extraction state. Now comparison is performed and if the compared feature is existed in the database, then the feature number is displayed; otherwise, miss match function is called.

4 Performance

Different numerical measurements can be analysed to measure the performance of a biometric system. The predominant performance attributes are FAR and FRR. The system begins the operation by individual enrolment, where the unique feature data depends on the kind of biometrics collection and storage in a feature database as template along with the user identification number. Then, the process of verification/validation to authenticate the biometric feature is done by the matching algorithm to decide either acceptance (or) rejection.

False Acceptance Rate The FAR is the possibility that the algorithm mistakenly permits a non-certified person, due to error template matching [11].

False Rejection Rate The FRR is the possibility that the application erroneously discards admission to a certified individual due to template mismatch [11].

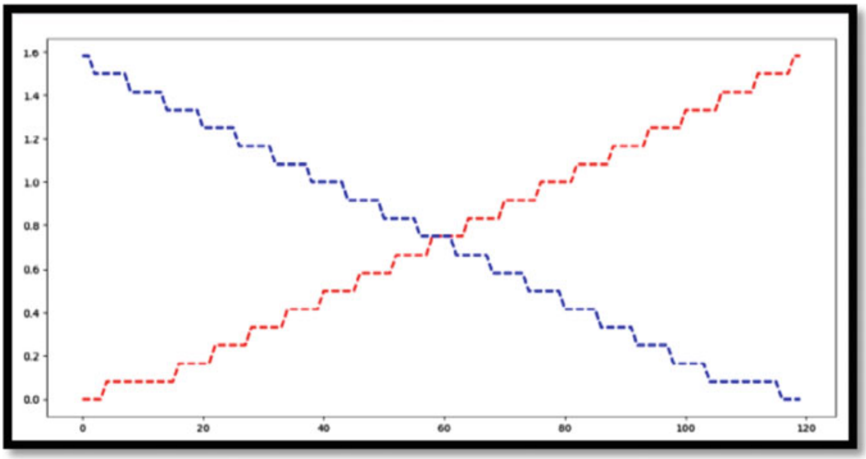


Fig. 3 Representation of FAR, FRR and CER

CER—Crossover Error Rate The CER is the representation of FAR and FRR curve, and it is the measurements of FAR and FRR intersection point.

FER—Failure to Enrol Rate The proportion of individuals who were unsuccessful to complete the registration. The reason for failure enrolment may be because of not familiar to the system, illumination (light) and surrounding environment.

Speed The time the biometric application device and verification/validation algorithm takes to enrol a sample and the time taken to verify/validate the individual subject.

Number of Templates In a general application, the highest quantity of patterns that can be stored is a concern. Most of the biometric tools are microcontroller-based systems where internal memory is limited. Therefore, the particulars of enrolment should also be limited.

Below is the performance calculation of biometric system with respect to FAR, FRR and CER. FAR and FRR values can be differentiated to maintain some threshold limit. If the value is below the threshold limit, then it is called FRR. If the threshold limit exceeds the imposters, then it is called FAR. A graph is plotted against FRR and FAR and the transaction point is determined as CER. Here both the FAR and FRR values are equal. The graphical representation of FRR and FAR with CER cross section point is shown. The red line represents FRR and the blue one represents FAR curve (Fig. 3) and Table 1 shows the extracted contour features.

Here, we compared three ear image database and calculated the performance measures (FAR, FRR, CER, speed). The databases are IIT Delhi, AMI and VR Students Sample databases with a total of 793, 699, and 100 ear images. The average time of execution is 0.33 s, and the average rate of FAR, FFR and CER is 0.8, 0.5 and 0.4%, respectively. These values are shown in Table 2.

Table 1 Extracted contour feature database

Image index	Aspect ratio	Perimeter	Area 1	*k*	*x*	*y*	*w*	*h*	Solidity	Extent	Equi_Diameter
0	1	5.6568542	2	True	17	6	3	3	0.5	0.22222222	1.595769122
1	2	11.656854	8	True	2	2	6	3	2	0.44444444	3.191538243
2	1	5.6568542	2	True	16	7	3	3	0.5	0.22222222	1.595769122
3	1	12.485281	10	False	16	6	5	5	2.5	0.4	3.568248232
4	1	5.6568542	2	True	13	6	3	3	0.5	0.22222222	1.595769122
5	1	11.899495	8.5	True	13	6	5	5	2.125	0.34	3.289762321
6	0.85	102.91169	177	False	1	1	29	34	44.25	0.17951319	15.01210843
7	2.8	40.142135	16	False	1	1	14	5	4	0.22857143	4.513516668
8	1	3.4142135	0.5	True	4	2	2	2	0.125	0.125	0.797884561
9	1.25	11.071068	8.5	True	6	4	5	4	2.125	0.425	3.289762321
10	1	5.6568542	2	True	28	14	3	3	0.5	0.22222222	1.595769122
11	0.75	7.6568542	4	True	29	15	3	4	1	0.33333333	2.256758334
12	1	5.6568542	2	True	28	14	3	3	0.5	0.22222222	1.595769122
13	0.75	7.6568542	4	True	38	32	3	4	1	0.33333333	2.256758334
14	1.85	33.213203	16.5	False	1	9	13	7	4.125	0.18131868	4.583497844
15	1.4	16.485281	9	False	1	8	7	5	2.25	0.25714286	3.385137501
16	2.5	33.899495	28.5	False	1	7	15	6	7.125	0.31666667	6.023896333
17	2	27.899495	18.5	False	1	8	12	6	4.625	0.25694444	4.85334231
18	2	2	0	False	9	15	2	1	0	0	0
19	1	6.2426406	1.5	True	9	13	3	3	0.375	0.16666667	1.381976598
20	1	6.2426406	1.5	True	9	13	3	3	0.375	0.16666667	1.381976598
21	0.5	2	0	False	18	88	1	2	0	0	0
22	1.66	9.6568542	1	False	14	88	5	3	0.25	0.06666667	1.128379167
23	1	6.8284271	2	False	12	89	3	3	0.5	0.22222222	1.595769122
24	2.2	50.627417	105	False	2	23	22	10	26.25	0.47727273	11.56244577
25	1	5.6568542	2	True	1	94	3	3	0.5	0.22222222	1.595769122

Table 2 Performance comparison of the three ear databases

Database name	# images in DB	Time of execution	FRR%	FAR%	CER%
IIT Delhi DB	793	0.339900	0.5	1.039	0.37
AMI DB	699	0.03500	0.5	1	0.455
VR Sample DB	100	0.05999	0.56	0.5	0.4

5 Conclusion

This work mainly focused on the contours to extract the features, such as bounding rectangle, aspect ratio, extent, equivalent diameter, checking the convexity, solidity, etc., to extract the contour properties of an ear image, which can be used to authenticate the individual on identity management systems. Here, IIT Delhi, AMI and VR Sample Ear Databases consisting of 793, 699 and 100 images, respectively, are used for experimentation. The experimental results show that this approach gives better results when compared to the traditional methods. This algorithm takes less amount of time to execute, and FAR, FRR are very low when compared to previous traditional mechanisms discussed in literature survey [12–17].

References

1. Attarchi S, Faez K, Rafiei A (2008) A New Segmentation Approach for Ear Recognition. In: Blanc-Talon J, Bourennane S, Philips W, Popescu D, Scheunders P. (eds) Advanced Concepts for Intelligent Vision Systems. ACIVS 2008. Lecture Notes in Computer Science, vol 5259. Springer, Berlin, Heidelberg
2. El-Bakry HM, Mastorakis N (2009) Ear recognition by using neural networks. In: Proceedings of the 11 th International Conference on Mathematical methods and computational techniques in Electrical engineering (pp 770–804)
3. Omara I, Li F, Zhang H, Zuo W (2016) A novel geometric feature extraction method for ear recognition. Expert Syst Appl 65:127–135
4. Kumar PR, Dhenakaran SS (2017) Structural (Shape) Feature Extraction for Ear Biometric System. In: Lobiyal D, Mohapatra D, Nagar A, Sahoo M (eds) Proceedings of the International Conference on Signal, Networks, Computing, and Systems. Lecture Notes in Electrical Engineering, vol 395. Springer, New Delhi
5. Yan P, Bowyer KW (2007) Biometric recognition using 3D ear shape IEEE Transactions on pattern analysis and machine intelligence 29(8):1297–1308
6. Kumar VN, Srinivasan B (2012) Ear biometrics in human identification system. Int J Inf Technol Comput Sci 4:41–47
7. Yuan L, Mu Z, & Xu Z (2005) Using ear biometrics for personal recognition. In: Advances in Biometric Person Authentication, Springer, Berlin, Heidelberg pp 221–228
8. Marti-Puig P, Rodríguez S, De Paz JF, Reig-Bolaño R, Rubio MP, & Bajo J (2012). Stereo video surveillance multi-agent system: new solutions for human motion analysis. Journal of Mathematical Imaging and Vision 42(2–3):176–195
9. Hurley DJ, Nixon MS, Carter JN (2000) Automatic ear recognition by force field transformations. In: IEE colloquium on vision biometrics (Ref. No. 2000/018). IET

10. Contour properties and features available in Opencv: http://docs.opencv.org/3.2.0/d3/d05/tutorial_py_table_of_contents_contours.html
11. Performance of biometrics: http://www.biometric-solutions.com/performance-of-biometrics.html
12. Pflug A, & Busch C (2012) Ear biometrics: a survey of detection, feature extraction and recognition methods. IET biometrics 1(2):114–129
13. Abaza A, Ross A, Hebert C, Harrison, MAF, Nixon MS (2013) A survey on ear biometrics. ACM computing surveys (CSUR), 45(2):22
14. Castrillón-Santana M, Lorenzo-Navarro J, Hernández-Sosa D (2011) An study on ear detection and its applications to face detection. In Conference of the Spanish Association for Artificial Intelligence, Springer, Berlin, Heidelberg pp 313–322
15. Lammi H-K (2004) Ear biometrics. Department of Information Technology, Lappeenranta University of Technology, Laboratory Information Processing, Lappeenranta, Finland
16. Choras M (2007). Image feature extraction methods for ear biometrics--a survey. In: 6th International Conference on Computer Information Systems and Industrial Management Applications (CISIM'07) IEEE. pp 261–265
17. Hurley DJ, Arbab-Zavar B, Nixon MS (2007) The ear as a biometric. In: Jain A, Flynn P, Ross A (eds) Handbook of biometrics, Chapter 7, Springer US, pp 131–150

Defocus Map-Based Segmentation of Automotive Vehicles

Senthil Kumar Thangavel, Nirmala Rajendran and Karthikeyan Vaiapury

Abstract Defocus estimation plays a vital role in segmentation and computer vision applications. Most of the existing work uses defocus map for segmentation, matting, decolorization and salient region detection. In this paper, we propose to use both defocus map and grabcut using wavelet for reliable segmentation of the image. The result shows the comparative analysis between the bi-orthogonal and Haar function using wavelet, grabcut and defocus map. Experimental results show promising results, and hence, this algorithm can be used to obtain the defocus map of the scene.

1 Introduction

Defocus estimation plays a vital role in computer vision applications including depth estimation, image quality assessment, blurring the image and image refocusing, etc. [1]. It can also be used for segmentation, matting, decolorization and salient region detection, depth estimation [2–5]. Practically, extracting a depth from a single image is very difficult because to estimate the depth of a point local features alone are not sufficient, global features should also be considered [6] defocus map provides some clues for depth estimation [7] (Fig. 1).

S. K. Thangavel (✉)
Department of Computer Science and Engineering, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Coimbatore, India
e-mail: t_senthilkumar@cb.amrita.edu

N. Rajendran · K. Vaiapury
TCS Research and Innovation, Tata Consultancy Services, Chennai, India
e-mail: rajendran.nirmala@tcs.com

K. Vaiapury
e-mail: karthikeyan.vaiapury@tcs.com

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_144

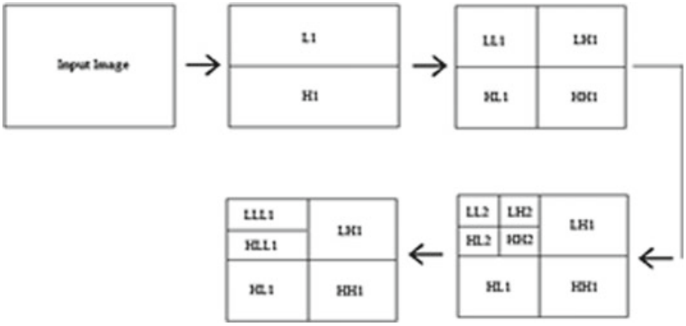


Fig. 1 Two levels of DWT decomposition

2D discrete wavelet transform (DWT) decomposition of the image can be represented [8, 9] as

$$C = X\ I\ Y \quad (1)$$

where C is the wavelet coefficients final matrix, I represents the input image and X and Y are the row filter matrix and column filter matrix, respectively. When DWT is applied on an image decomposition of the image takes place which decomposes the image into four parts in first level, namely, approximation coefficient consists of low-low frequencies, horizontal coefficient consists of low-high frequencies, vertical coefficient consists of high-low frequencies and diagonal coefficient consists of high-high frequencies. The approximation coefficient computed in the first level is used for the next level of decomposition.

To reconstruct the image, inverse discrete wavelet transform (IDWT) is used which can be represented as

$$I = X^{-1}\,C\,Y^{-1} \quad (2)$$

for the matrices which are orthogonal Eq. (2) can be modified [8] as

$$I = X^{T}\,C\,Y^{T} \quad (3)$$

There are different types of wavelet families, namely, Haar wavelet, bio-orthogonal wavelet, Daubechies wavelet and coiflets wavelets [9].

Although considerable work has been done by the research community on defocus map, we propose to use both defocus map and grabcut using wavelet for segmenting the image. We also provide a comparative analysis between the bi-orthogonal and Haar function using wavelet, grabcut and defocus map.

In this work, we have used both defocus map and grabcut using wavelet for segmenting the image. The wavelet families used in these works are Haar and bi-orthogonal function.

2 Related Works

Defocus map is used to highlight the most important subpart of the object. There are various approaches that are used for defocus estimation like active illumination method (AIM), coded aperture method (CAM), interpolation method (IM), deconvolution, etc. [10, 11]. The defocus map can also be obtained from multiple images, but the major problems that can be faced are occlusions.

Although considerable work has been done by the research community on defocus map, we propose to use both defocus map and grabcut using wavelet for segmenting the image. We also provide a comparative analysis between the bi-orthogonal and Haar function using wavelet, grabcut and defocus map.

In [7], Zhuo et al. proposed a method for finding the depth of a single image. The blur estimation method using Gaussian gradient ratio (GGR) is used to compute the depth. Initially, the defocus blur at the edges in an input image is calculated using step function. Then, the point spread function which is also known as Gaussian kernel function is convoluted with the edge pixels. The output obtained with the isotropic Gaussian kernel is re-blurred. GGR is computed between the original blur image and the re-blur image to get the maximum edge location. This results in the sparse defocus map. Due to the inaccurate blur at the edge locations, joint bilateral filtering applied to the sparse defocus map to improve the efficiency. The sparse defocus map is applied in the entire image to get the depth map; this can be achieved by using defocus blur interpolation. Matting Laplacian is used for defocus blur interpolation and the depth map is obtained. This approach is robust to noise, inaccurate blur and the adjacent neighbour edges. The limitation of this approach is that it cannot provide any information, whether the blur at the edges are obtained due to blur or defocus in the input image.

In [2], Jiang et al. proposed an algorithm for salient region detection, which uses three cues such as uniqueness, focus and objectness. In their work, the focus/defocus estimation is done by using scale-space analysis; objectness estimation is computed using window overlapping-based approach. The uniqueness is estimated using pixel level and regional level uniqueness. In [4], Namboodiri et al. proposed a method to estimate the depth from a single image using reverse heat equation and depth from defocus (DFD) using graph cuts. In their proposed method, experiments are performed on various types of images, and the results show that relative layers of depth can be estimated.

In [8], Petrova et al. proposed a method for detecting the edges in medical images using wavelets. In their proposed work they have done detailed experiments on different types of edge detection methods that can be done using wavelets like modification of approximation replaced by zero, modification of approximation using edge detectors and wavelet transform modulus maxima method. In [9], Brannock et al. proposed a method for detecting the edges using a modification of approximation replaced by zero. In their work, the compared the different types of wavelet families like Haar wavelet, bi-orthogonal wavelet, Daubechies wavelet and coiflets wavelets for edge detection on noisy images.

Grabcut segmentation is used to segment the 2D image. By drawing a rectangle in an image, the region of interest is marked. The user needs to segment the image as foreground and background regions with respect to the region of interest.

Gaussian mixture models are used to modal the foreground and the background. Grabcut is the combination of statistical modal and graph cuts of the background and foreground region.

The paper is organized as follows: In the first section, we have introduced defocus map and discrete wavelet transforms. Section 2 provides the related works. In Sect. 3, the proposed solution and description are given. Section 4 provides the defocus map estimation, and Sect. 5 gives a detail description of the results and conclusion. Finally, the conclusion and future work are given in Sect. 6.

3 Proposed Solution

The given input image is converted to monochrome image which consists of black and white image. Discrete wavelet transform is computed on the monochrome image which decomposes the image into four parts such as approximation coefficient, horizontal coefficient, vertical coefficient and diagonal coefficient. The approximation coefficient is alone used for further processing. This work explains various types of segmentation using grabcut and approximation coefficients. The experiments are done for two wavelet families, namely, Haar and bi-orthogonal function.

A. Modification of approximation replaced by zero

Approximation coefficient obtained by DWT is replaced with zero, which removes all the low-level frequencies and the results obtained is superimposed with the monochrome image and the edges are extracted (Figs. 2 and 3).

B. Modification of approximation replaced by zero using grabcut segmentation

Grabcut segmentation is used to segment the monochrome image as foreground and background regions. The foreground region consists of objects of interest and DWT is computed on the resultant output. The approximation coefficient obtained is replaced with the zeros, and the edges are extracted by superimposing zero approximation coefficient with the monochrome image.

Fig. 2 Block diagram for defocus map using modification of approximation replaced by zero

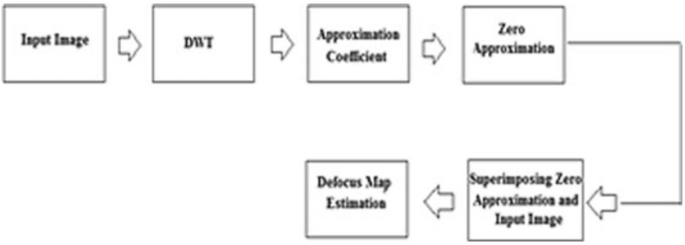


Fig. 3 Block diagram for defocus map using modification of approximation by edge projection

C. **Modification of approximation by edge detection**

By using approximation coefficient obtained by DWT, the edge pixels are detected using edge detectors. Then, the resultant image is superimposed with the monochrome image.

D. **Modification of approximation by projection using grabcut segmentation**

The foreground and the background regions are segmented using grabcut segmentation to get the region of interest. DWT is computed on the foreground region and the approximation coefficient is obtained. The edge pixel is detected on the approximation coefficient. The resultant image is superimposed with the monochrome image.

4 Defocus Map Estimation

This approach has been divided into two phases (Fig. 4):

1. Sparse defocus blur estimation and
2. Defocus blur propagation.

Initially, the gradient-based method is used to compute the defocus blur on the edges. Then, the sparse defocus map is propagated to the whole image. In sparse defocus map when the input image $I(m, n)$ is given, the edges are extracted using edge detection algorithms. Then, the defocus blur is calculated by assuming the edge

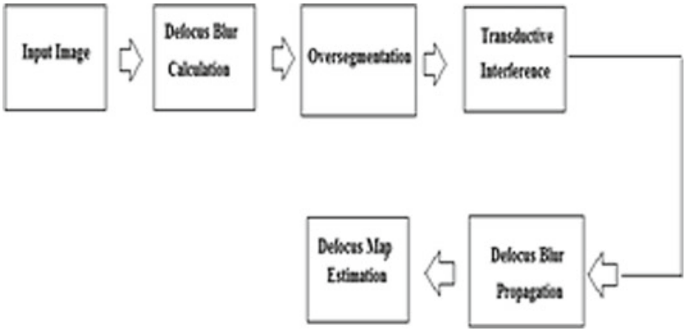


Fig. 4 Block diagram for defocus map estimation

function as $x(p) = af(p) + b$, where a and b are the amplitude and offset of the edge for image $I(m, n)$. $x(.)$ indicates the step function and p be the pixel location of the image $I(m, n)$. The edge function $x(p)$ and Gaussian kernel $G_k\ (p, s)$ are convoluted to obtain the defocus blur and s is the unknown standard deviation and the degree of blurness present on the edges. The defocus blur can be written as

$$D_f(x(p)) = x(p) * G_k(p, s) \tag{4}$$

Another Gaussian kernel is used to re-blur the

$$\nabla\big(D_f(x(p)) * G_k(p, s_1)\big) = \nabla x(p) * G_k(p, s) * G_k(p, s_1)$$

$$= \frac{a}{\sqrt{2\pi\left(s^2 + s_1^2\right)}} \exp -\left(\frac{p}{s^2 + s_1^2}\right) \tag{5}$$

where s_1 is the standard deviation of re-blur Gaussian kernel. Gradient magnitude ratio (G_M) between the original blur and the re-blur gives the maximum value at the edges, which is denoted as

$$G_M = \frac{D_f(x(p)}{(D_f(x(p) * G_{k1}(p, s_1)} \tag{6}$$

$$G_M = \sqrt{\frac{s^2 + s_1^2}{s^2}} \tag{7}$$

$$s = \frac{s_1}{\sqrt{G_m^2 - 1}} \tag{8}$$

using gradient magnitude ratio the unknown standard deviation are calculated. The over-segmentation is used to find the super-pixels in the input image which can be used to find the similarity between the adjacent pixels. Simple linear iterative clustering (SLIC) is used to find the super-pixels in the given input image using k means clustering [6]. The super-pixels can be represented as $M = \{M_1, M_2 \ldots M_i\}$. In the super-pixels set M, the weighted is added to the graph $W_{\text{cg}} = (M, X, \alpha)$ where α varies from [0, 1], M is the super-pixel set and X is the edge set which contains the all two adjacent super-pixel. The weight matrix can be defined as

$$W = \left[\alpha_{ij}\right]_{n\times n} \tag{9}$$

Using over-segmentation, similarity between the adjacent pixels alone can be calculated. To calculate the similarity between any super-pixels which is not just limited to the adjacent pixel, transductive matrix is used. This is obtained by using the affine matrix which can be written as

$$\begin{gathered} T_i = (D - \varphi\alpha)^{-1} I \\ \varphi = (0, 1) \end{gathered} \tag{10}$$

The initial blur function P_{mj} for each super-pixel is can be represented as

$$P_{mj} = \underset{p \in X_j}{\text{med}} \, T(P) \tag{11}$$

The defocus blur propagation using affine information can be written as

$$\widetilde{P_{mj}} = \widetilde{T_i}[P_{m1}, P_{m2} \ldots P_{mn}] \tag{12}$$

5 Results and Discussion

Figures 5, 6, 7, 8, 9, 10, 11 and 12 represent the outputs obtained using edge pixel projection using bi-orthogonal and Haar wavelet functions. The comparison has been made by using bi-orthogonal and Haar functions using discrete wavelet transform with or without grabcut segmentation. Figure 5a represents the results for superimposed modification of approximation replaced by zero with original image by using bi-orthogonal function, and Fig. 5b represents the defocus map obtained for Fig. 5a; the edges are extracted but are not much clear, and the outliers are also extracted.

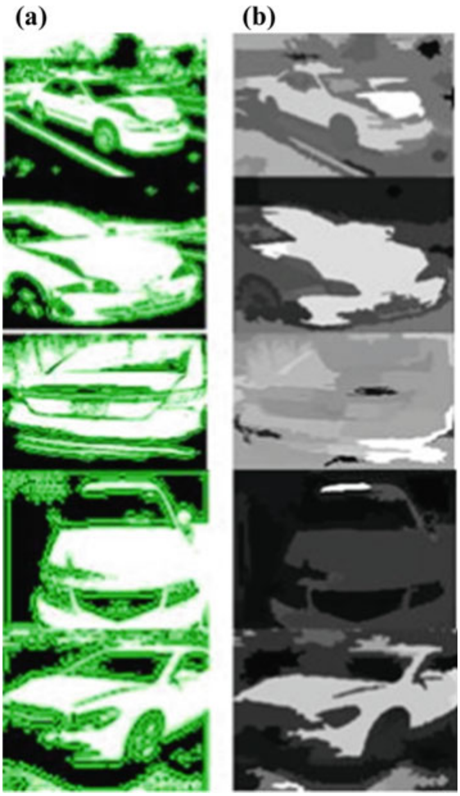


Fig. 5 **a** Superimposed original image and zero approximation for bi-orthogonal wavelet function. **b** Defocus map for (**a**)

Figure 6a shows the results by using superimposed modification for the approximation by zero and original image with grabcut segmentation for using bi-orthogonal function from which the edges are extracted. Figure 6b shows the defocus map results for Fig. 6a which gives the better results compared to modification of approximation method by zeros using bi-orthogonal wavelet function.

Figure 7a shows the results for superimposed modification of approximation replaced by edge projection with original image using bi-orthogonal wavelet function and Fig. 7a defocus map results obtained for Fig. 7a.

In Fig. 8a, the results obtained for superimposed modification of approximation using edge with original image and grabcut segmentation using bi-orthogonal wavelet function is provided. Figure 8b shows the results for defocus map of Fig. 8a. Comparing the results obtained using bi-orthogonal function, Fig. 8 gives better segmentation result since outliers are reduced in a significant manner. In Fig. 9a, the results for superimposed modification of approximation replaced by zero with original image by using Haar wavelet function are provided. Figure 9b represents the defocus map obtained for Fig. 9a. The edges are extracted and the defocus map obtained is not clear. Figure 10a represents the results for superimposed modification of approxima-

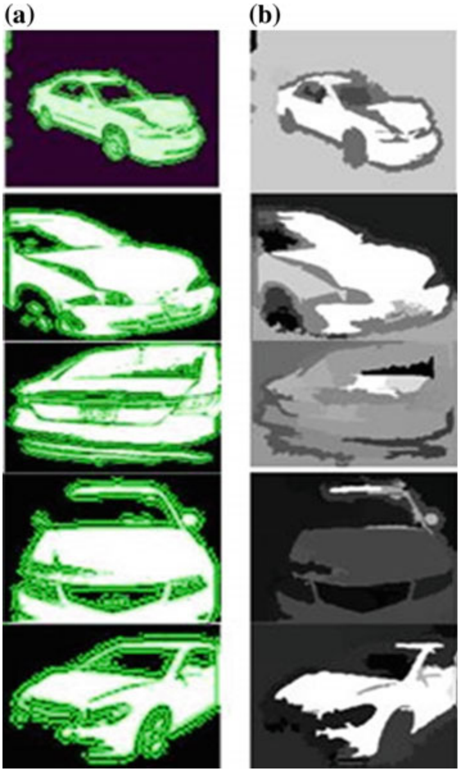


Fig. 6 **a** Superimposing of original image and zero approximation using grab cut for bi-orthogonal wavelet function. **b** Defocus map for (**a**)

tion replaced by zero with original image and grabcut segmentation by using Haar wavelet function and Fig. 10b represents the defocus map obtained for Fig. 10a. The edges are extracted and defocus map obtained is better compared to modification of approximation replaced by zero with original image using Haar wavelet function.

Figure 11a represents the results for superimposed modification of approximation by edge projection with the original image by using Haar wavelet function and Fig. 11b represents the defocus map obtained for Fig. 11a.

In Fig. 12a, the results for superimposed modification of approximation replaced by edge projection with original image and grabcut segmentation by using Haar wavelet function is provided. Figure 12b represents the defocus map obtained for Fig. 12a. The edges are extracted and the defocus map obtained is better compared to modification of approximation replaced by zero with original image using Haar wavelet function. By comparing the results obtained using both bi-orthogonal wavelet function and Haar wavelet function, Haar wavelet using DWT in Fig. 12 gives better results.

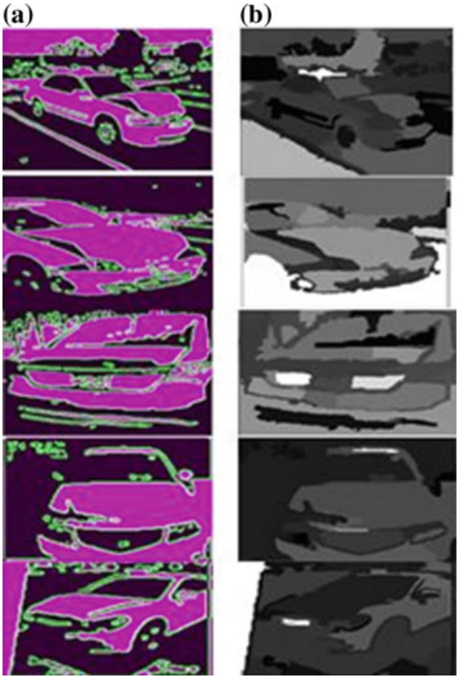


Fig. 7 **a** Superimposing of original image and edge projection for bi-orthogonal wavelet function. **b** Defocus map for (**a**)

Fig. 8 **a** Superimposing of original image and edge projection using grabcut for bi-orthogonal wavelet function. **b** Defocus map for (**a**)

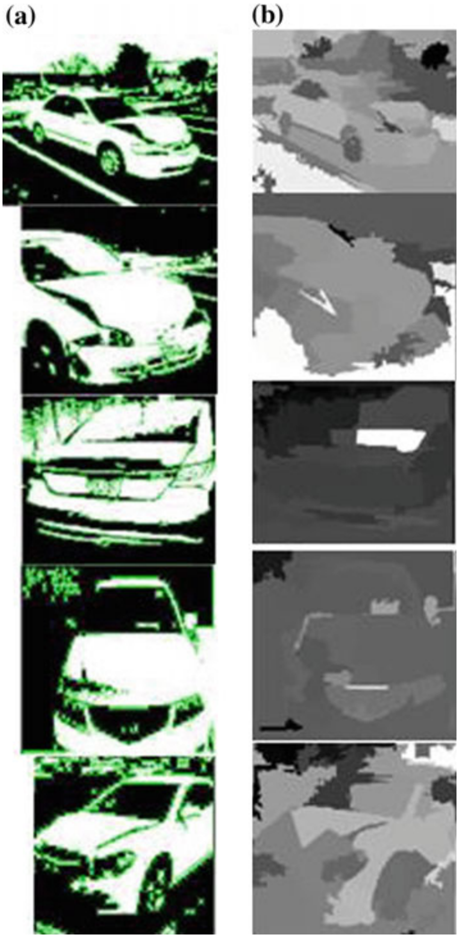


Fig. 9 **a** Superimposing of original image and zero approximation for Haar wavelet function. **b** Defocus map for (**a**)

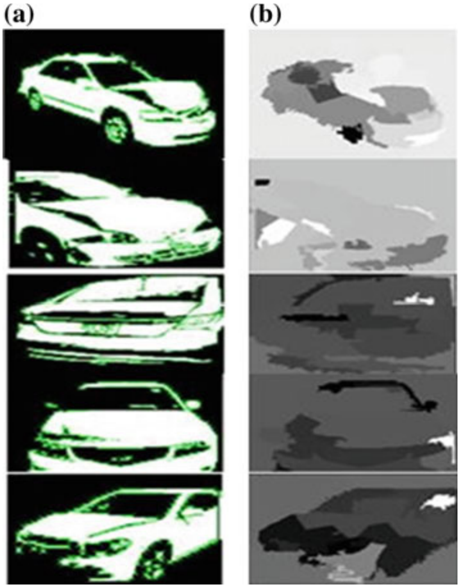


Fig. 10 **a** Superimposing of original image and zero approximation using grabcut for Haar wavelet function. **b** Defocus map for (**a**)

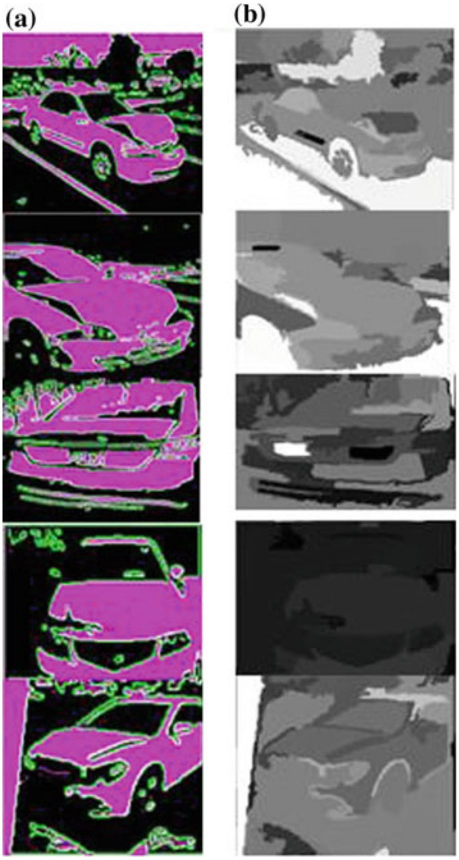


Fig. 11 **a** Superimposing of original image and edge projection for Haar wavelet function. **b** Defocus map for (**a**)

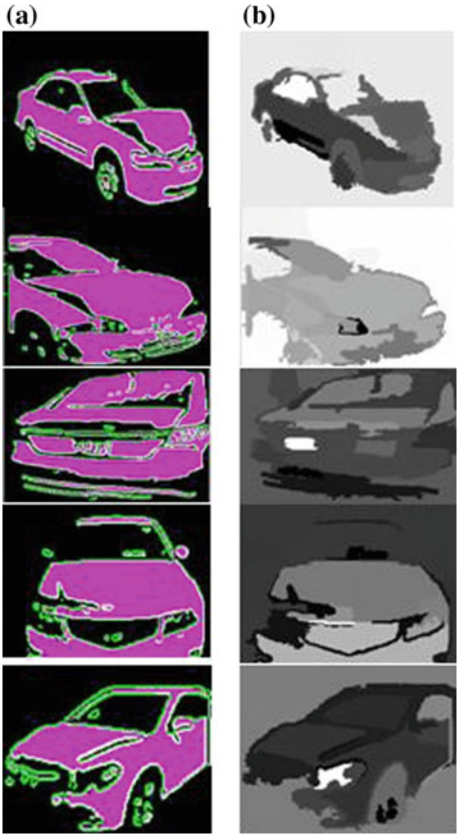


Fig. 12 **a** Superimposing of original image and edge projection using grabcut for Haar wavelet function. **b** Defocus map for (**a**)

6 Conclusion and Future Work

In this work, we have proposed a solution to use both defocus map and grabcut using wavelet for segmenting the image. The comparative analyses have been made between the bi-orthogonal and Haar function using wavelet, grabcut and defocus map.

Acknowledgements This work is performed as a part of internship with TCS Innovation Lab.

References

1. Chen D-J, Chen H-T, Chang L-W (2016) Fast defocus map estimation. In: 2016 IEEE international conference on image processing (ICIP). IEEE
2. Jiang P, Ling H, Yu J, Peng J (2013) Salient region detection by UFO: uniqueness, focusness and objectness. In: IEEE international conference on computer vision, ICCV 2013, Sydney, Australia, 1–8 Dec 2013, pp 1976–1983
3. Lin HT, Tai Y, Brown MS (2011) Motion regularization for matting motion blurred objects. IEEE Trans Pattern Anal Mach Intell 33(11):2329–2336
4. Namboodiri VP, Chaudhuri S (2008) Recovery of relative depth from a single observation using an uncalibrated (real-aperture) camera. In: 2008 IEEE computer society conference on computer vision and pattern recognition (CVPR 2008), 24–26 June 2008, Anchorage, Alaska, USA
5. Peng Y, Zhao X, Cosman PC (2015) Single underwater image enhancement using depth estimation based on blurriness. In: 2015 IEEE international conference on image processing, ICIP 2015, Quebec City, QC, Canada, 27–30 Sept 2015, pp 4952–4956
6. Ashutosh S, Chung SH, Ng AY (2005) Learning depth from single monocular images. NIPS 18:1–8
7. Zhuo S, Sim T (2011) Defocus map estimation from a single image. Pattern Recogn 44(9):1852–1858
8. Petrova J, Hostalkova E (2011) Edge detection in medical image using the wavelet transform. Report of research, Department of Computing and Control Engineering, Czech Public
9. Brannock, E, Weeks M (2006) Edge detection using wavelets. In: Proceedings of the 44th annual Southeast regional conference, pp 649–654. ACM
10. Moreno-Noguer F, Belhumeur PN, Nayar SK (2007) Active refocusing of images and videos. ACM Trans Graph 26(3):67–75
11. Levin A, Fergus R, Durand F, Freeman WT (2007) Image and depth from a conventional camera with a coded aperture. ACM Trans Graph 26(3):70–78

Multi-insight Monocular Vision System Using a Refractive Projection Model

J. Mohamed Asharudeen and Senthil Kumar Thangavel

Abstract The depth information of a scene, imaged from the inside of a patient's body, is a difficult task using a monocular vision system. A multi-perception vision system idea has been proposed as a solution in this work. The vision system of the camera has been altered with the refractive projection model. The developed lens model recognises the scene with multiple perceptions. The motion parallax is observed under the different lenses for the single shot, captured through the monocular vision system. The presence of multiple lenses refracts the light in the scene at the different angles. Eventually, the appearance of the object dimension is augmented with more spatial cues that help in capturing 3D information in a single shot. The affine transformations between the lenses have been estimated to calibrate the multi-insight monocular vision system. The geometrical model of the refractive projection is proposed. The multi-insight lens plays a significant role in spatial user interaction.

1 Introduction

A multi-insight vision is built to observe the world with a single camera system. The stereo vision system is used to observe the shape, size and localization of the target from two or more images taken at different views. The proposed work aims to retrieve the target information with the single camera system. It handles distinct perceptions with a single shot. The new lens model is constructed with five distinct lenses. The parent lens of the model holds the normal view. Its neighbourhood lens holds the information hidden in the parent lens. The designed model of the lens is termed as augmenting the refraction (ATR lens). From here, augmenting the refraction lens has

J. Mohamed Asharudeen · S. K. Thangavel (✉)
Department of Computer Science and Engineering, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Coimbatore, India
e-mail: t_senthilkumar@cb.amrita.edu

J. Mohamed Asharudeen
e-mail: cb.en.p2cvi16002@cb.students.amrita.edu

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_145

been termed as ATR lens. The motion parallax is the key to the proposed lens model observed in the neighbourhood lenses.

The proposed method takes advantage of the refraction provided by the lens. The image captured using the ATR lens holds the refracted ray of the neighbourhood lenses.

The depth estimation can be done using visual cues from the images. The popular algorithm used for recovering the depth information such as stereopsis, structure from motion uses multiple images. The features extracted from a single image are not sufficient enough to gather the depth information. Global features, as well as local features, are considered for extracting the depth information [1]. The multi-insight image that contains visual cues of the same scene at the different angles gives the required depth information. The acquisition of spatial cues with ATR is more prominent. The existing work perceives the monocular cues from a single image. The proposed work acquires the visual cues from a single image. Here, the image is a multi-insight image and thus holds more spatial cues information. As an initial prototype lens model, it is made bigger in size.

This paper is arranged as follows: the background works made in calibration and a few works about the projection model are discussed. The consecutive section holds the proposed ATR lens model and then the extended pinhole model of the camera. The geometric model of the light rays traced in the ATR lens is interpreted. The estimation of the target using the proposed geometry model is observed in the later section.

2 Related Works

Cui et al. [2] presented a perspective projection model for the stereovision system. It considers the prism as a single lens. It establishes an affine transformations relation between the image points and the object points. Beardsley and Murray [3] presented the calibration method by making use of the vanishing line and vanishing point information.

Zhang [4] proposed a flexible new technique for calibrating a camera. This requires a planar pattern. The results are better without using any kind of orthogonal plates or the distinct world points. Maximum likelihood estimation is used for the refinement. Teixeira et al. [5] implemented the algorithm proposed by [4]. This gives the understanding of Zhang's method. It estimates the intrinsic, extrinsic and radial distortion parameters. Burger [6] gives a detailed report about the steps involved in Zhang's calibration method. Cipolla et al. [7] presented a method for recovering the projection matrices from uncalibrated images to reconstruct the 3D models. The geometrical intuitive method has been interpreted using parallelism and orthogonality of the architectural scenes.

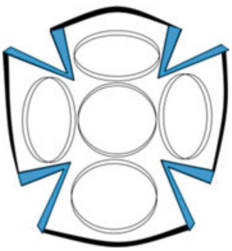
Tsai [8] introduced a new 3D camera calibration technique for machine vision technology with the advantage in terms of accuracy, speed rather than the existing state of the art. Jurjevic and Gasparovic [9] presented a two-step calibration technique and 3D point cloud estimation with open source technologies. This gives the general introduction of epipolar geometry. The distortion coefficients are also discussed in detail. Heikkila and Silven [10] presented a simple technique of calibrating a camera with four steps. The linear estimation and nonlinear estimation of projection matrix are also discussed.

Barath et al. [11] developed a new method of estimating the focal length and the fundamental matrix with the help of two affine correspondences from the uncalibrated cameras. Meng and Hu [12] presented a new calibration pattern for calibrating a camera. The calibration technique is closely related to the existing Zhang's method. Feng et al. [13] presented a new transparent calibration board for calibrating cameras at multiple angles. They introduced two more new parameters such as depth and refractive index of the glass. Still, it uses the traditional calibration method. Drennan [14] gives a general idea of the camera calibration algorithm. The calibration of a camera using vanishing points and vanishing lines are also discussed in [15–18]. Michels et al. [1] used a hierarchical multiscale Markov random field to recover the depth information. The depth model is estimated using global and local features present in a single image. The visual cues are discussed in comparison with the binocular cues.

3 Proposed Works

In this section, the ATR model is elaborated. The refractive projection model has been constructed using five distinct lenses. The lenses are arranged in a fashion as shown in Fig. 1. These lenses perceive the environment at different angles as shown in Fig. 2 to form the images. These images are responsible for observing the motion parallax, it is termed as the multi-insight image. A single camera vision system is responsible for the novel vision.

Fig. 1 Lens arrangement



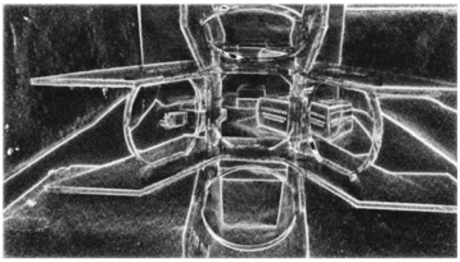


Fig. 2 ATR lens prototype

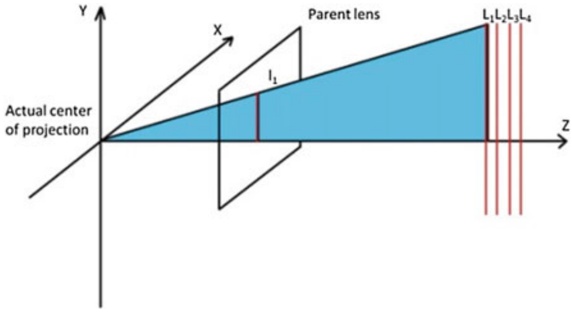


Fig. 3 Pinhole projection model

3.1 Camera Model

Image formation made by the ATR lens is described in this section. The image formation of a pinhole camera model shown in Fig. 3 is altered for the ATR lens model. The projection of a world point in space onto the image plane follows the similar triangle principle [19–21]. The centre of projection [22] in the Euclidean coordinate system is represented as COP. Let the world point be X and its image point be x. A new virtual centre of projection (VCOP) is introduced shown in Fig. 4 from which the refracted rays meet the actual COP. The VCOP is assumed for calibrating the individual lens. The motion parallax plays a significant role in ATR lens. Objects that are nearer moves faster than the faraway object which has been distinguished with its distinct lens. The occluded surfaces from the parent lens are less obscure in the neighbourhood lens.

Every individual lens holds the similar triangle property for mapping the world coordinates $(X, Y, Z)^T$ to the image coordinates. The proposed system claims that the camera coordinate frame has been replaced with lens coordinate frame as the system involves more than one lens. The neighbourhood lens still follows the similar triangle rule as parent lens. The VCOP of the neighbourhood is shifted to the actual centre of projection.

The mapping for the parent lens is from the Euclidean space R^3 to R^2. The mapping for the neighbourhood lens is also from the Euclidean space R^3 to R^2, and then from R^2 to R^2.

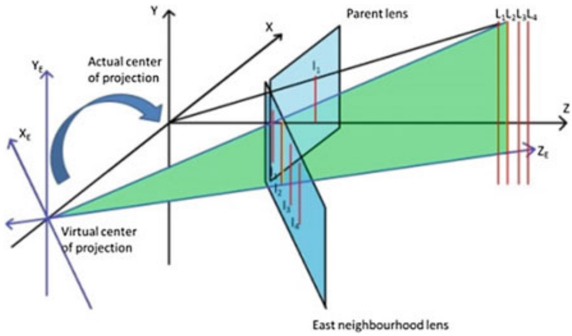


Fig. 4 Extended pinhole model of ATR lens

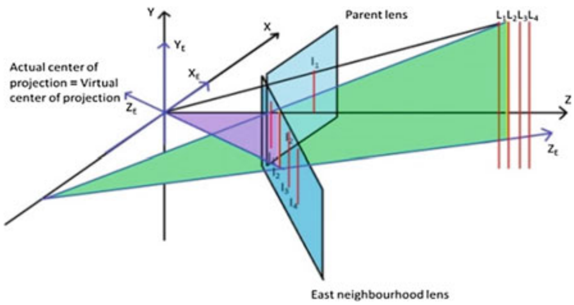


Fig. 5 Virtual centre of projection shifted to the actual centre of projection

The linear mapping between the homogeneous coordinates of the world coordinate system and the image coordinate system is represented as in Eq. (1). Let x_a be the homogeneous vector of the image coordinates that correspond to the actual centre of projection, and X be the homogeneous vector of the world coordinates. Let P be the camera projection matrix.

$$x_a = PX \tag{1}$$

The neighbourhood lens follows the linear mapping similar to the parent lens between the homogeneous coordinates of the world coordinate system and the lens coordinate system as in Eq. (2), and then, it follows Eq. (3) for the linear mapping between the homogeneous coordinates of the lens coordinate system, and the image coordinate system corresponds to the actual centre of projection. Let x_L be the homogeneous vector of the lens coordinates. Let P_R be the refractive projection matrix. The refractive projection matrix represents the shifting of the neighbourhood coordinates from the VCOP to the actual centre of projection with a rotation and a translation vectors.

$$x_L = PX \tag{2}$$

$$x_a = P_R x_L \tag{3}$$

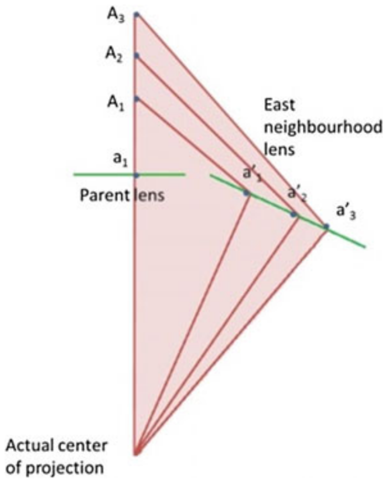


Fig. 6 Ray converges to the actual centre of projection in ATR model

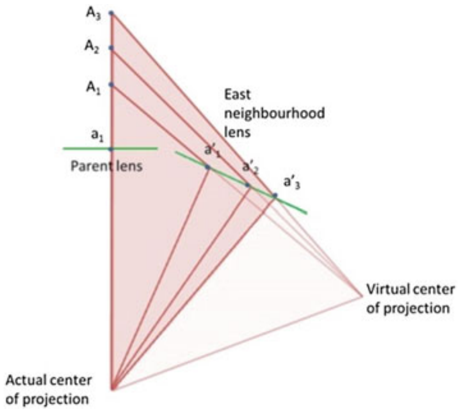


Fig. 7 Ray visualised to converge at the virtual centre of projection

The ray perception for the ATR lens is shown in Fig. 6. The parent lens sets the neighbourhood lens as an observer to retrieve the occluded information. The horizontal section of the prototype is examined instead of the entire set-up. Three lenses are placed in the set-up. The refracted rays from the neighbourhood lenses are observed in the multi-insight image. Let's consider target $A3$ is occluded by $A2$ and is occluded by $A1$. These targets are observed through neighbourhood lens. VCOP of the right neighbourhood lens is shown in Fig. 7. The VCOP is shifted to the actual centre of projection. Once it is shifted, the ray observation is traced as shown in Fig. 5.

The entire ray observation of the horizontal section is shown in Fig. 8. The occluded targets are observed by the neighbourhood lenses.

The affine transformation obeys the Thales theorem as shown in Figs. 9 and 10. The normal to the neighbourhood lens meets at a point k. The distance between the point k and the actual centre of projection of parent lens is maintained between its neighbourhood lenses to their virtual centre of projection.

According to Thales theorem, point k, centre of the lens and corresponding projection centre follow the theorem as in Eq. (4). Thus, the orientation can be estimated easily and is fixed.

$$\Delta\, ADE \text{ where } BC||DE$$

$$\frac{AB}{BD} = \frac{AC}{CE} \tag{4}$$

The intrinsic parameter estimated from the parent lens is utilised as the intrinsic parameter for the neighbourhood lens, as the calibration parameter estimated depends on the single camera [23, 24]. The affine transformation matrix is constructed with rotation only along the y-axis for the horizontal section and along the x-axis for the vertical section of the lens. Translation is from the VCOP and the actual centre of projection. The ATR lens set-up works to find out the occluded parts in the scene with multiple perceptions. The estimation of the target follows the proposed occluded triangles relations. The model for target estimation is shown in Fig. 11. The vertex of the rays viewed as a series of the triangle with a common vertex and a common base. It has the other two vertexes that are collinear. The occluded triangles are constructed as right-angled triangles with every vertex of all triangles are collinear, and they are having neither common base nor common vertex.

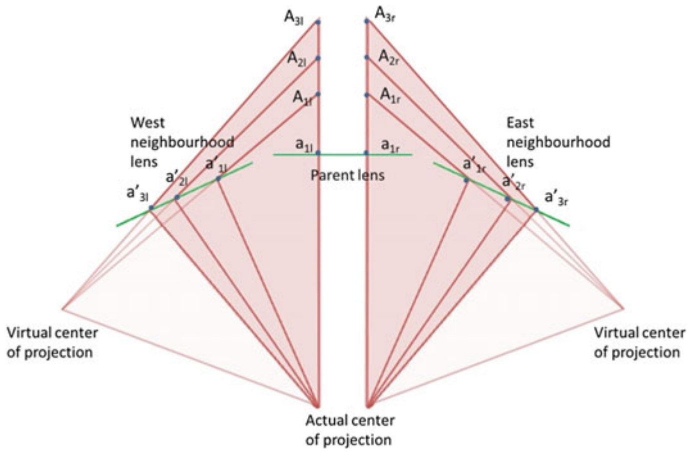


Fig. 8 Ray convergences of the entire ATR model

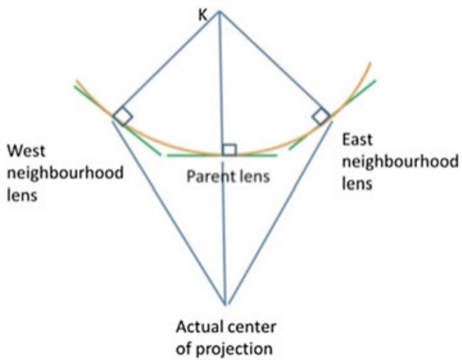


Fig. 9 Actual centre of projection

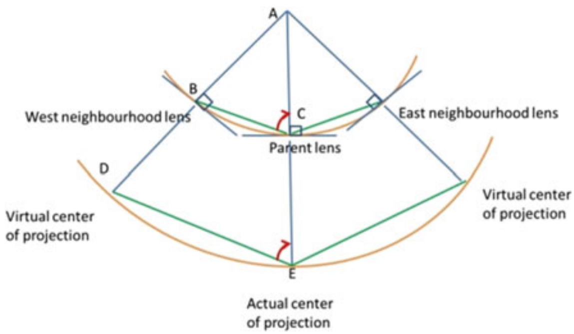


Fig. 10 Estimation of virtual centre of projection

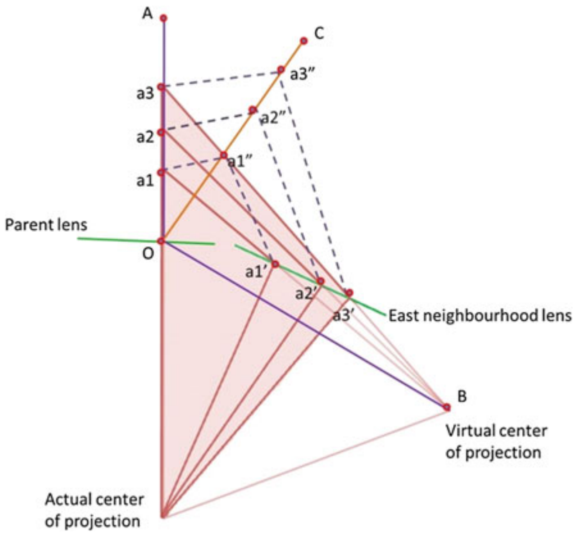


Fig. 11 Target estimation using geometrical model

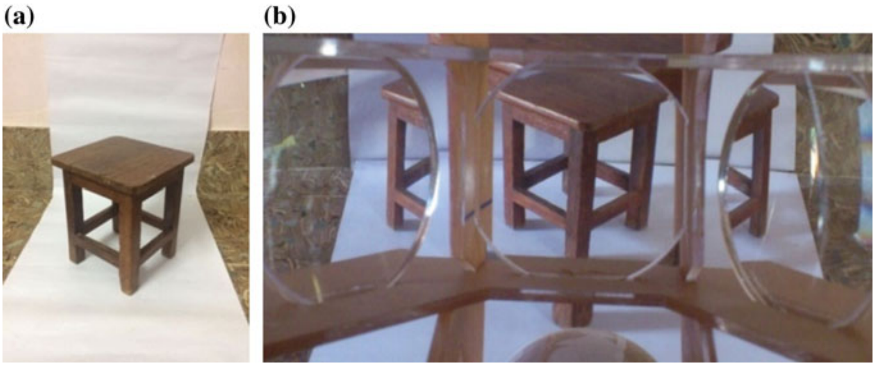


Fig. 12 **a** Image captured in a camera and **b** multi-insight image

4 Results

The ATR lens is developed with an intention to be utilised in medical imaging vision system. As it is a prototype model and needed lens calibration, this model is developed in bigger size for making the calibration easier to explore the multi-insight images. The image captured through the lens model consists of simple targets. These targets are traced using feature detectors. The image captured using a single camera is shown in Fig. 12a. The dimension of the target captured with different perceptions in multi-insight image is shown in Fig. 12b. The comparison of the image is shown in Fig. 12 to understand the visual cues are observed only at this single image about the target. The visual cues of the target are observed with different perceptions in this multi-insight image.

The image captured within the lens region gets segmented and is shown in Fig. 13a–c from west neighbourhood lens, from parent lens and east neighbourhood lens, respectively.

These segmented images are shown in Fig. 14 and have the feature mapping between them [25–28]. From Fig. 14, it is evident that the depth information from the multi-insight image can be estimated once the correspondence matching has been made. The calibration of the lens is made using Zhang's calibration method.

5 Conclusion

The refractive projective model has been modelled as an extended pinhole model. The monocular vision system developed with this model handles distinct perceptions when imaged from the inside of the patient's body. The multi-insight image obtained with this ATR lens holds more cue information about the scene. The motion parallax plays a significant role in ATR lens. The visual cues obtained from multi-insight

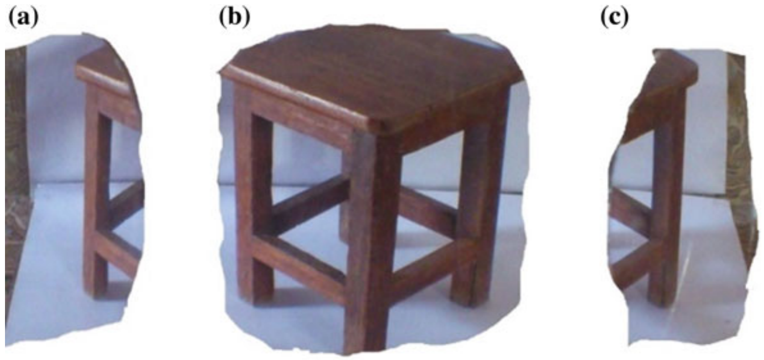


Fig. 13 Image from the lens is segmented from the multi-insight image: **a** segmented from west neighbourhood lens, **b** segmented from parent lens and **c** segmented from east neighbourhood lens

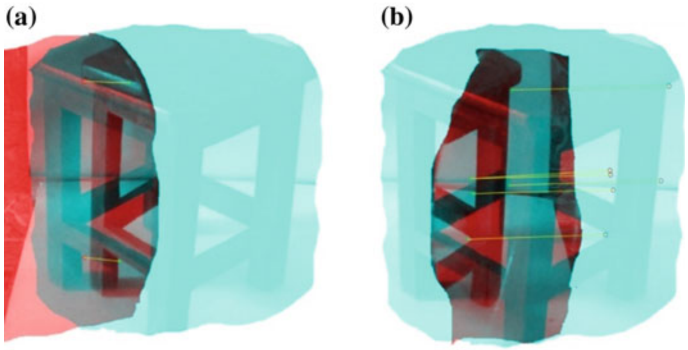


Fig. 14 Features matched with the parent lens image and the neighbourhood lens image: **a** mapping between parent lens and west neighbourhood lens image, and **b** mapping between parent lens and east neighbourhood lens image

images can be utilised for constructing the depth information. The multi-insight lens plays a significant role in spatial user interaction.

References

1. Michels J, Saxena A, Ng AY (2007) 3D depth reconstruction from a single still image. Int J Comput Vis
2. Cui X, Zhao Y, Lim K, Wu T (2015) Perspective projection model for prism based stereo vision. Opt Soc Am 23(21)
3. Beardsley P, Murray D (1992) Camera calibration using vanishing points. University of Oxford, Oxford
4. Zhang Z (1998) A flexible new technique for camera calibration. Microsoft research-technical report-98-71

5. Teixeira L, Gattass M, Fernandez M (2006) Zhang's camera calibration: step by step
6. Burger W (2016) Zhang's camera calibration algorithm: in-depth tutorial and implementation. Technical report-HGB16-05
7. Cipolla R, Drummond T, Robertson D (1999) Camera calibration from vanishing points in images of architectural scenes. University of Cambridge, BMVC99
8. Tsai RY (1987) A versatile camera calibration technique for high accuracy 3D machine vision metrology using off the shelf TV cameras and lenses. IEEE J Robot Autom RA-3(4)
9. Jurjevic L, Gasparovic M (2017) 3D data acquisition based on OpenCV for close range photogrammetry applications. Int Arch Photogram Remote Sens Spat Inf Sci XLII-1/W1
10. Heikkila J, Silven O (1997) A four step camera calibration procedure with implicit image correction. University of Oulu
11. Barath D, Toth T, Hajder L (2017) A minimal solution for two view focal length estimation using two affine correspondences. Machine Perception Research Laboratory
12. Meng X, Hu Z (2003) A new easy camera calibration technique based on circular points. J Pattern Recogn Soc: 115–1164
13. Feng H, Li Q, Feng M (2017) A novel multi camera calibration method using transparent glass calibration board. In: Proceedings of 2nd international conference on advances in materials machinery, Electronics I. American Institute of Physics
14. Drennan M (2010) An implementation of camera calibration algorithms. Clemson University
15. Orghidan R, Salvi J, Gordan M, Orza B (2012) Camera calibration using two or three vanishing points. In: IEEE proceedings of the federated conference on computer science and information systems, pp 123–130
16. Wang L-L, Tsai W-H (1990) Computing camera parameters using vanishing line information from a rectangular parallelepiped. Mach Vis Appl 129–141
17. Guillou E, Meneveaux D, Maisel E, Bouatouch K (2000) Using vanishing points for camera calibration and coarse 3D reconstruction from a single image. Vis Comput 396–410
18. Caprile B, Torre V (1990) Using vanishing points for camera calibration. Int J Comput Vis 4:127–140
19. Hartley R, Zisserman A (2005) Multiple view geometry in computer vision. Cambridge University Press, Cambridge
20. Beardsley P, Murray D, Zisserman (1992) A Camera calibration using multiple images. University of Oxford, Oxford
21. Feng M, Jia X, Wang J, Feng S, Zheng T (2017) Global calibration of multi camera based on refractive projection and ray tracing. Sensors 17:2494
22. Hartley RI, Kaucic R (2002) Sensitivity of calibration to principal point position. In: European conference on computer vision. Springer, Berlin
23. Rothwell CA, Zisserman A, Marinos CI, Forsyth DA, Mundy JL (1992) Relative motion and pose from arbitrary plane curves. University of Oxford, Oxford
24. Hansen P, Alismail H, Rander P, Browning B (2012) Online continuous stereo extrinsic parameter estimation. CVPR
25. Sreelakshmi S, Vijai A, Senthil Kumar T (2016) Detection and segmentation of cluttered objects from texture cluttered scene. In: Proceedings of the international conference on soft computing systems
26. Senthil Kumar T, Vijai A (2014) 3D reconstruction of face: a comparison of marching cube and improved marching cube algorithm. Int J Adv Image Process Tech 1:6–9
27. Nirmala R, Thangavel SK (2017) Develop, implement and evaluate a multimodal system for government organisation. In: Proceedings of 4th international conference on advanced computing and communication systems, ICACCS 2017, pp 111–117
28. Kavin D, Thangavel SK (2017) Improving capabilities for visually challenged person in library environment. In: Proceedings of 4th international conference on advanced computing and communication systems, ICACCS 2017, pp 1–10

ONESTOP: A Tool for Performing Generic Operations with Visual Support

Gowtham Ganesan, Subikshaa Senthilkumar and Senthil Kumar Thangavel

Abstract Programming has become tedious for every person these days. Learning programming languages and writing a computer program for different tasks using various programming languages is a difficult and time-consuming task. Therefore, modules are used to make programming easier and faster. Cloud computing enables applications to be accessed everywhere. The 'ONESTOP' tool will be provided as a facility to the users under the category 'Software as a Service'. The paper provides directions for enabling the same facility. It does not address the challenges for provisioning this tool on the cloud. Every module in ONESTOP consists of the operations under that category. The tool processes the input by removing fillers, identifying the operation to be performed using trie data structure and synonym mapping and displaying the result. User need not write codes or define functions. A simple sentence in English is sufficient to perform the task. The tool is easy to use and does not require any programming knowledge to use it. All the operations are performed in less time enhancing the performance of the tool. Key aspect of ONESTOP is that it does not produce any error and saves debugging time.

1 Introduction

In recent times, people assume that programming knowledge is the only need for building an application. But, for a proper structuring of a program, one has to plan the steps required to solve the problem. Then, structuring the program is important. To achieve this, modules are used. It helps programmers reuse modules thereby making

G. Ganesan · S. Senthilkumar · S. K. Thangavel (✉)
Department of Computer Science and Engineering,
Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Coimbatore, India
e-mail: t_senthilkumar@cb.amrita.edu

G. Ganesan
e-mail: cb.en.u4cse14417@cb.students.amrita.edu

S. Senthilkumar
e-mail: cb.en.u4cse14451@cb.students.amrita.edu

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_146

the task easier. And, novice programmers will find it easy to use a module or function rather than writing it again.

Fidge [1] has proved in his experiment that the common programming mistakes were caused because of lack of experience in programming in a particular language, not because of the way of thinking programmatically. Hence, the time spent on learning the syntax of a computer language can instead be effectively spent to develop an algorithm or to focus more on other important aspects. ONESTOP is mainly developed to encourage people to explore and create new algorithms rather than concentrating on learning programming languages. The purpose of the tool is to provide the result of a given pseudocode so that users may spend time doing useful work. This will help in enhancing the problem-solving skills of people.

Students are exposed to various operations in a small period of time. They tend to forget it or learn its result without proper understanding. Practical knowledge always helps us remember things for a longer time than compared to theoretical knowledge. ONESTOP provides a platform for students to perform operations and view the results so that they will always remember it. Unique feature of this tool is that users need not enter the operations to be performed in the form of a code. User has to specify a query in English to perform an operation and the tool will provide the result. By doing so, it saves time and user will get a visual support of how the result will look like. And, user can select operations from the list of operations in the tool and enter input values instead of typing the whole input query.

Tokenization is an important method in identifying keywords from a sentence. It plays a major role in understanding the actual meaning of a sentence by breaking it into tokens. After splitting a sentence into tokens, tokens are analysed about their significance in the sentence. ONESTOP uses tokenization and other techniques to generate the result. The tool is developed in C++, thereby ensuring less execution time. Qt Creator is used to maintain cross-platform consistency. ONESTOP is designed meticulously that there is no possibility for an error to occur.

The paper is organized as follows: Section 2 gives the reason why the tool has been developed. In Sect. 3, various works are analysed and their limitations are discussed. Section 4 describes how the tool works. Pseudocode, screenshots of the result, comparison of similar tools and user experiments are discussed in Sect. 5. Section 6 comprises future work and Sect. 7 contains conclusion.

2 Motivation

People believe that programming skills are required to do any task. This forces people to learn various programming languages related to their task or field. They will also have to remember the syntax of every language. This can be time-consuming. Many tools have been developed in recent days which aim to reduce the need to remember syntax of various languages. ONESTOP is developed to perform operations without the need to code. Even a person without programming knowledge will be able to use this tool at ease and perform the necessary operation. ONESTOP aims to prove that programming knowledge is not always necessary to perform an operation. Even a basic sentence can be used to perform a task. Hence, people can focus on learning new algorithms or explore new topics instead of spending time on learning programming languages.

3 Related Work

Charntaweekhun and Wangsiripitak [2] proposed a system for visual programming using flowchart. In their system, when a flowchart is drawn, it is compiled and run and the output is displayed on the system environment. The main aim of this project is that novice programmers do not have to write codes. They can draw many flowcharts, link them all and when the run button is pressed, a source code in C is generated and EXE file is invoked to display the output. Similarly, ONESTOP aims to remove the burden of writing codes faced by programmers. Programmers can give commands in a form of sentence in English instead of using flowcharts to get the output. Also, giving commands in English would be much easier than drawing flowcharts.

Sarda and Jain [3] have eliminated the need to know query language to retrieve data from database. And user need not know the database schema. The system accepts any form of query which is given as a set of words, and then the words are translated into database queries with the help of metadata. The result is then provided to the user. Disadvantage of this tool is that it does not understand the purpose and usage of quotation marks in queries.

Little and Miller [4] understand the difficulty of programming using syntax. So they came up with an idea of syntax-free programming wherein a parser would generate code for the given input in the form of English words. ONESTOP provides the result and description instead of generating codes. This is because novice programmers would have a better understanding of the input provided when results are provided with description rather than providing codes.

Smith et al. [5] were concerned that all the emails could not be read or categorized since large number of emails sent by Anne Hunter, Course VI administrator are going around. So the team developed a project to categorize emails and to generate keywords and key phrases for every mail so that students will know whether the email is important or not. In this project, Tynan and others have used regular expression pattern for classification from email texts. ONESTOP also used regex patterns for classifying if input belongs to arithmetic or complex number type.

Fudaba et al. [6] have developed a tool called Pseudogen. As the name suggests, the tool generates pseudocode in English or Japanese from source code written in Python. This tool was developed since people had difficulty in understanding source codes that are written in programming languages which are not well known. People can further make use of ONESTOP to know the result and working of instructions which are provided in English.

Tahir and Ahmed [7] proposed a Tree-Combined Trie or TC-Trie which is a combination of trees and tries for fast IP address lookup. ONESTOP also uses trie data structure for prefix matching to identify the operations to be performed under a category. This trie-based search operation happens after analysing the input query.

Senthil and Ohhm [8] developed a queueing model for e-learning system. Since the tool is stored in cloud, if many people are trying to access the tool at the same time, this queueing model can be used. Sankar and others [9] conducted a survey on different machine learning algorithms. Naive Bayes algorithm was discussed and its insights are used in user experiment of this tool.

4 Methodology

In this section, the working of the tool is discussed. Thereby, the step-by-step processing of the tool will be understood.

The tool is stored in cloud so that many people can access it. There are two types of cloud users who can access the tool. (i) Guest users—only few operations like arithmetic and trigonometric operations under number operations category will be available to these users. (ii) Privileged users—these users should log in using username and password. They have to pay to use the tool. All the operations present in the tool such as the operations under number, string, matrix, complex number category can be used by privileged users.

Figure 1 explains the block diagram of the tool. Initially, the user provides a sentence in English. The sentence can be used to perform arithmetic, relational, complex, matrix, geometric and trigonometric operations. The sentence or set of words is subjected to two refining mechanisms to minimize the sentence length and aids in reducing its processing complexity. The first mechanism of input sentence reduction is to remove the articles. It is removed as it is unnecessary and does not provide any useful meaning to the sentence or for further processing. The second mechanism is to remove white spaces in the sentence. This is done as the tokens are stored in dictionary without spaces between words. White spaces are also removed in order to eliminate storing of similar tokens in the dictionary. The given sentence is converted to lower case since all the tokens are stored in lower case in the dictionary. Dictionary refers to the words used in the tool. If some characters are in lower case and some in upper case, searching those tokens in dictionary will be difficult as all possible combinations of lower and upper case words should be stored in dictionary. So, for easy searching of tokens, all the words in the given input are converted to lower case. While removing white spaces and converting the sentence to lower case, there is an exception to do these processes. The exception is when single or double quotes are used in the query. The data within the quotes are not modified and kept as such.

After processing the input, the tool checks if the input is of arithmetic or complex pattern. The pattern is a regular expression. When the input follows a definite format, regular expression is the ideal one to be used. If operations to be performed are given as a word (e.g. Add), then we find what operation it has to perform by using trie-based search. In case the operation to be performed is given in the form

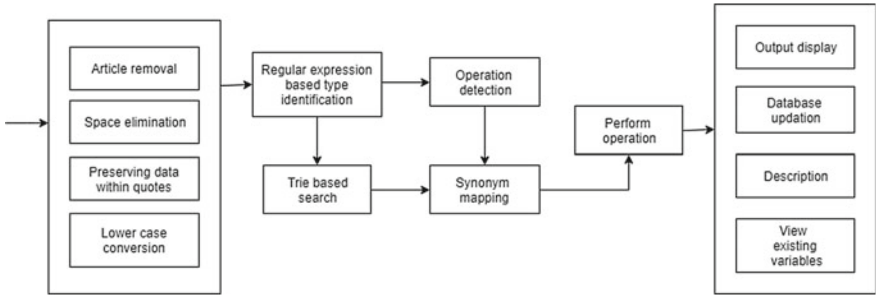


Fig. 1 Block diagram

of a symbol (e.g. +), we need not identify what operation should be performed by searching the trie. Instead, the operation will be found out with the help of regular expression patterns. In this tool, we have used two regular expression patterns: (i) [0–9] * [\+\−][1–9][0–9] * *i* is the pattern for complex numbers and (ii) ([0–9]|[\+\−\%\(\)]) + is the pattern used to identify arithmetic operations. In case the user specifies the wrong input, the compiler will throw an error and user can specify the right input.

If the processed input matches the arithmetic regular expression pattern, the tool identifies which arithmetic operation has to be performed and implements it. On the other hand, when the input matches the complex number pattern, tool knows that it has to perform operations in accordance to complex number category. But the tool will not know what operation should be performed (i.e.) if it has to perform addition or subtraction, etc. To help this out, ONESTOP uses trie-based search to find out the operation.

Trie belongs to tree data structure. Trie is used for searching the string efficiently. Every node in trie has index value and Boolean variable. Boolean variable is used to indicate that the right word is identified by traversing the trie until end of the word. To make the string searching efficient, trie data structure is used in this tool. By using trie, complexity of string searching has been reduced from $O(n^3)$ to $O(n^2)$. This thereby enhances the computational efficiency of the tool. When contents present in the trie is large, it gives rise to higher space complexity. This is a disadvantage of trie. In the proposed tool, we have overcome this disadvantage by inserting the possible operations of a particular category (string, arithmetic, complex, matrix, etc.) into the trie. We repeat this process of inserting the possible operations into the trie for all categories and searching is performed simultaneously. After searching the trie, we will know the categories (e.g. String, arithmetic) in which a particular operation (e.g. add) can be performed. Based on the values given in the input (for instance, either string values or numbers) to perform an operation, the tool will know which category (addition of strings in case of string and basic addition in case of numbers) will be appropriate since operations will be performed for values of same data type. If the values do not belong to a uniform data type, the tool generates an alert box and confirms with the user on basis of what data type, the operation should be performed.

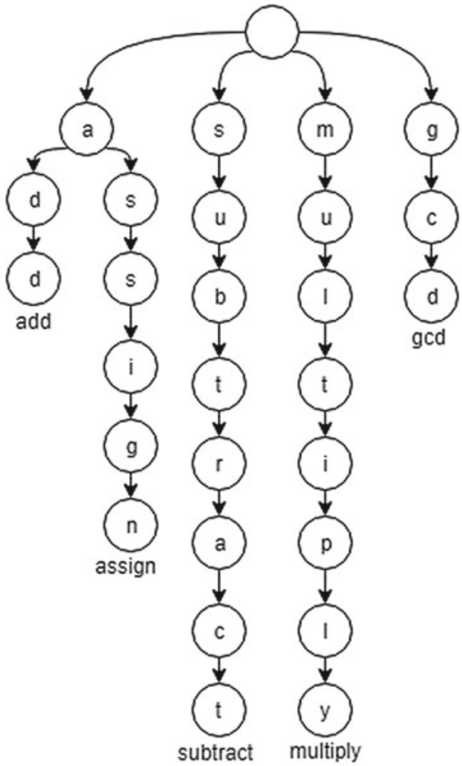


Fig. 2 Trie data structure

To perform search operation in trie, the tool traverses the query first. The possible operations in a particular category (say Matrix) are inserted into the trie. The tool checks if the first character in sentence matches the character in the root node of trie. If they match, next character in the input is checked with the child of the root node in trie and this process repeats for every character in the input which is checked with the subsequent child nodes until the Boolean variable in a node is encountered, search operation stops for that category and returns false. The tool then searches all other categories. In the end, we will get an array of possible categories (e.g. matrix, arithmetic, complex) which contains the operation (e.g. add) found from the input.

Figure 2 shows the example of trie data structure for the operations under the matrix category. They include add, subtract, assign, multiply and gcd. The root node is null. Trie data structure is based on prefix matching. Since prefix matching is used, common prefix of two or more words is stored only once in trie. Add, assign operations in Fig. 2 explain this. To search for multiply in the above trie, the first letter (*m*) is checked. Then, the consecutive letters are checked. On checking the last letter (*y*) in the word (multiply), the word is found and true is returned.

Map data structure is used in the proposed tool to associate similar tokens with the main token. Main tokens are assigned by the tool developer. For instance, sum, total, aggregate (similar tokens with same meaning) are mapped to add (main token). Operations are performed based on main tokens only. If suppose similar tokens are not mapped to the main token, every similar token has to be assigned the same operation individually. This will require a lot of memory space and will lead to unnecessary checking of all similar tokens. Since mapping is done in this tool, all similar tokens are collectively mapped to one main token. By doing so, the operation identification process is done in an efficient manner. Expanded words are also mapped to their corresponding ACRONYM. For instance, Total Surface Area is mapped to TSA.

The operation that has to be performed is found so far. From this, the tool will find out the minimum number of operands required to execute the operation by classifying them as unary, binary and ternary operators. Performing the operation using the operands from the input sentence comes under three ways (i) the tool will extract the values given in the input query, (ii) if previously existing variables are used in the input, the tool will check the database for the value of that variable and retrieve it and (iii) if the number of values or previously existing variables specified in the input sentence do not meet the minimum number of required values to perform the operation, prompt box will be generated and user will be asked to enter the values. From the values obtained, operations are computed, and the result is displayed to the user.

When the result is provided to the user, the result and the variables used for computation are stored in the database. The textual description provides details about how the variables are obtained. The textual description also includes how the result was obtained. Along with the result, the variables which contain the values used to perform the operation are also displayed. Thereby, user will know the value present in every variable which is used to perform the operation. User can make of these previously existing variables later by mentioning them in the input.

Another important feature of this tool is that user need not specify the operation in form of a sentence or set of words. There are inbuilt options in the tool and the user just has to select what operation he/she wants to perform. After selecting the operation to be done, user can enter the values in the prompt box or select the previously existing variables. This aims to make the tool user-friendly and reduce the time taken to enter input sentences.

5 Discussion

Pseudocode

Function: Search (input, array of words)

- Insert all the array elements received as parameter into trie data structure.
- For i = 0 to length(input):

(i) Search the input sub-string starting from 'i' in trie
(ii) If match is found in trie, break the loop

- Return the found item

(Time Complexity: O (n^2)
Space complexity: O (domain size being searched))

Steps:

1. Enter the input sentence.
2. Pre-processing the input

- Initialise an array 'articles' containing 'a', 'an', 'the'.
- Remove the words present in the article array from input sentence.
- Eliminate spaces between words in the input(only for words outside single and double quotes)
- Convert input to lower case(only for characters outside single and double quotes)

3. If previously existing variables are used in input,

then
data types of those variables are found from the database containing information about variables, data types and values.

4. Check the input if it matches regular expression pattern

- if input is of complex operation pattern,

then
assign operation type as 'complex'.

- else if input is of arithmetic pattern,

then
assign operation type as 'arithmetic'
and perform the operation.

- Else if the names of categories of operations are specified in the input,

then
Search (input, category names) can be invoked to identify the category by passing the category names as parameter to the trie.

5. If category of operations are identified,

then

- Assign all the operations available in that category to a temporary variable.
- Invoke Search (input, temporary variable).
- Obtain the returned item.

6. Else if category of operations is not found,

then

- Initialize an array 'numtypes' with four values 'number', 'string', 'complex' and 'matrix'
- For i = 0 to length(numtypes):

i. Assign all the operations available for numtypes[i] to a temporary variable
ii. Invoke Search (input, temporary variable).
iii. Add the returned operations of (ii) to an array.

- If values required are specified in input,

then
Analyse the input values to determine the category (as string, number or complex) from (iii).

- If category cannot be found from input value,

then
user will be asked to select the desired category from (iii).

- Assign the operation found to 'operation' variable.
- A map variable 'otherwords' contains key value pairs where root words are given as keys.
- Map words with same meaning to root word.
- A map variable 'operands' where keys are operation names and values are minimum number of operands required.
- Analyse input sentence in three steps to get input values to perform the operation

i. append the values mentioned in the input to 'input' array
ii. values of existing variables specified in the input are added to 'input' array.
iii. if length('input' array) < operands[operation] :
then
obtain the values from user.

- Operation is performed and result is stored.

5.1 Result

Figure 3 shows input given by user, result along with description and the variables—input1, input2 and result1 are stored in the database. In this case, to calculate the total surface area of a cone, the required values are provided in the input query itself. Here, the word total surface area is mapped to TSA and value of TSA of a cone is calculated.

Figure 4 shows how the TSA of a cone is calculated using previously existing variable. Result1 contains the value got by performing the TSA calculation from input values given in Fig. 3. The value of slant height is provided by the user as 5, TSA of cone is calculated and result is provided.

In case the user just specifies what operation should be formed and does not mention any value or previously existing variable in the input query, the tool generates a prompt asking for the necessary values to perform the operation. In Fig. 5, to find TSA of a cone, user has not specified any value. So, the tool will generate a prompt box as in Fig. 6 asking for the required input (radius). Another prompt box will be generated by the tool asking the user to provide value of slant height. Based on these values, TSA of a cone is computed and result is displayed.

To perform operations using matrices, if the user provides an input query as given in Fig. 7, prompt box pops up requesting the number of matrices to be multiplied (Fig. 8). Then two more prompt boxes will pop up asking for the number of rows and columns for each matrix

User then has to enter values for each matrix as in Fig. 9 and the result, variables used and description will be displayed to the user.

If the user wants to perform complex number operations such as addition of two complex numbers as in Fig. 10, given input is checked if it matches the regular expression pattern for complex numbers and 'add' is found out to be the operation to be formed as a result of trie-based search. After this process, the real and imaginary parts of two complex numbers are added and result is displayed.

The tool is developed in a user-friendly manner wherein the user has to select what operation he/she wants to perform under a category (Fig. 11) and then enter the required values. This will reduce the time taken to enter the query by the user and make the work easy.

5.2 Similar Tools

The proposed tool has all the features which are limitations of the tools mentioned in Table 1. In ONESTOP, input sentence can be provided for complex logic also. Instead of drawing complex flowcharts as in Flowgorithm, which are difficult to understand, sentences can be specified in a sequential manner for easy understanding. Raptor enforces the user to follow predefined structure whereas ONESTOP does not expect the user to follow any structure to specify the input. To use ONESTOP, user does

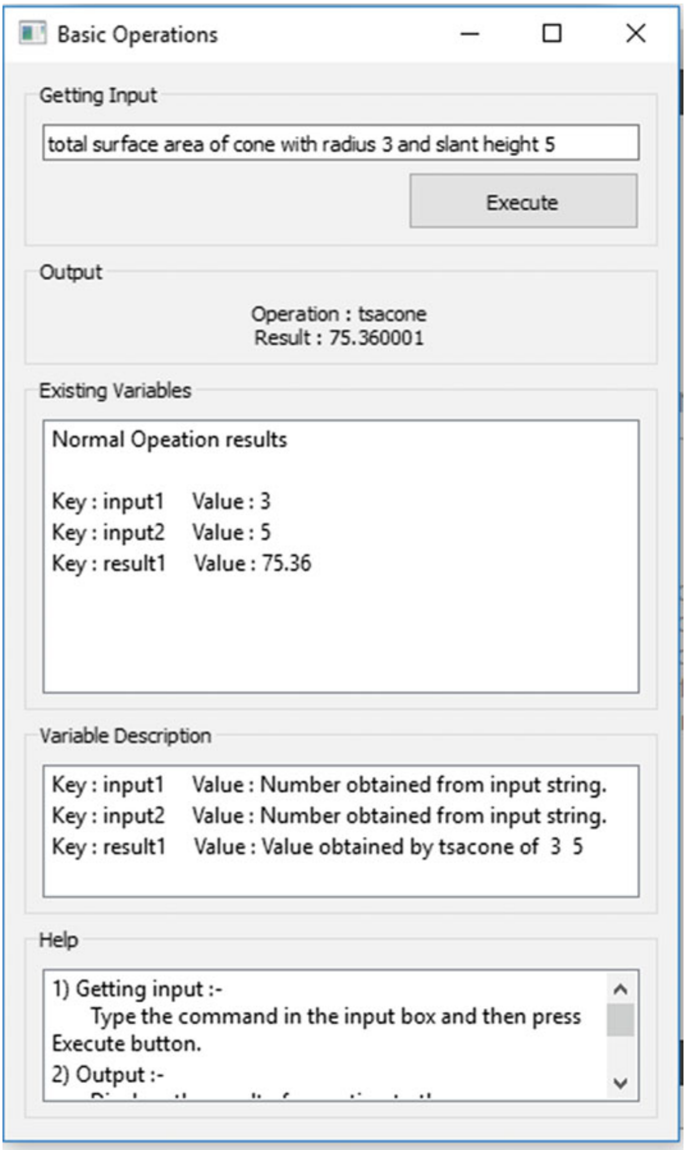


Fig. 3 Total surface area performed with all values specified

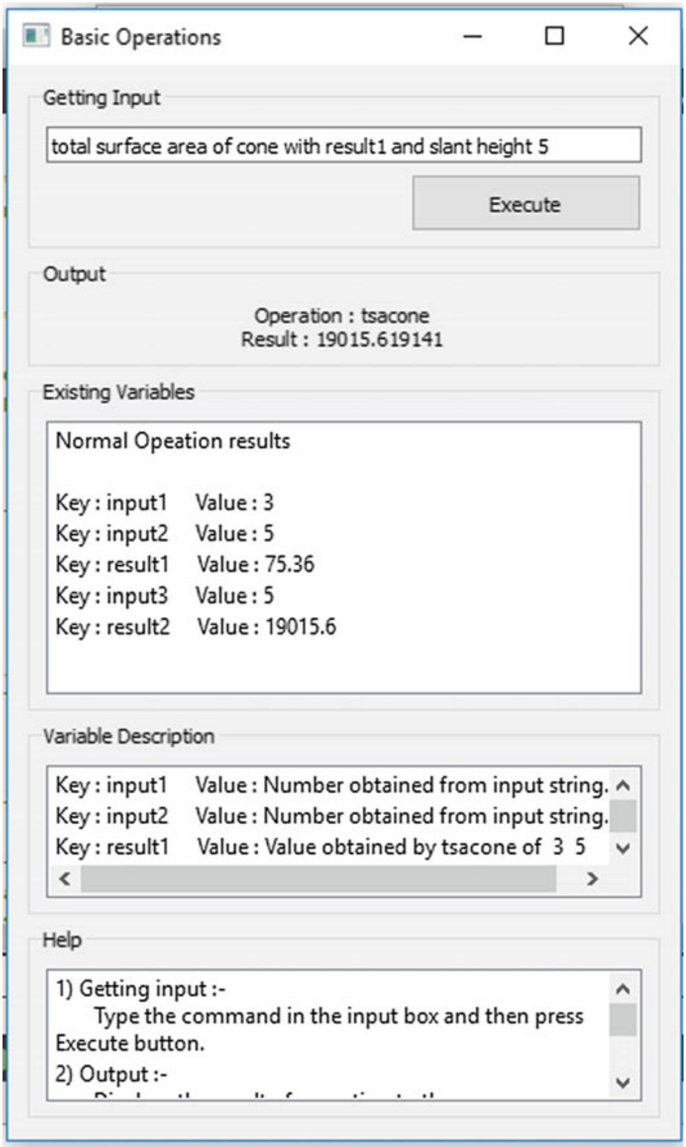


Fig. 4 Total surface area performed using existing variable

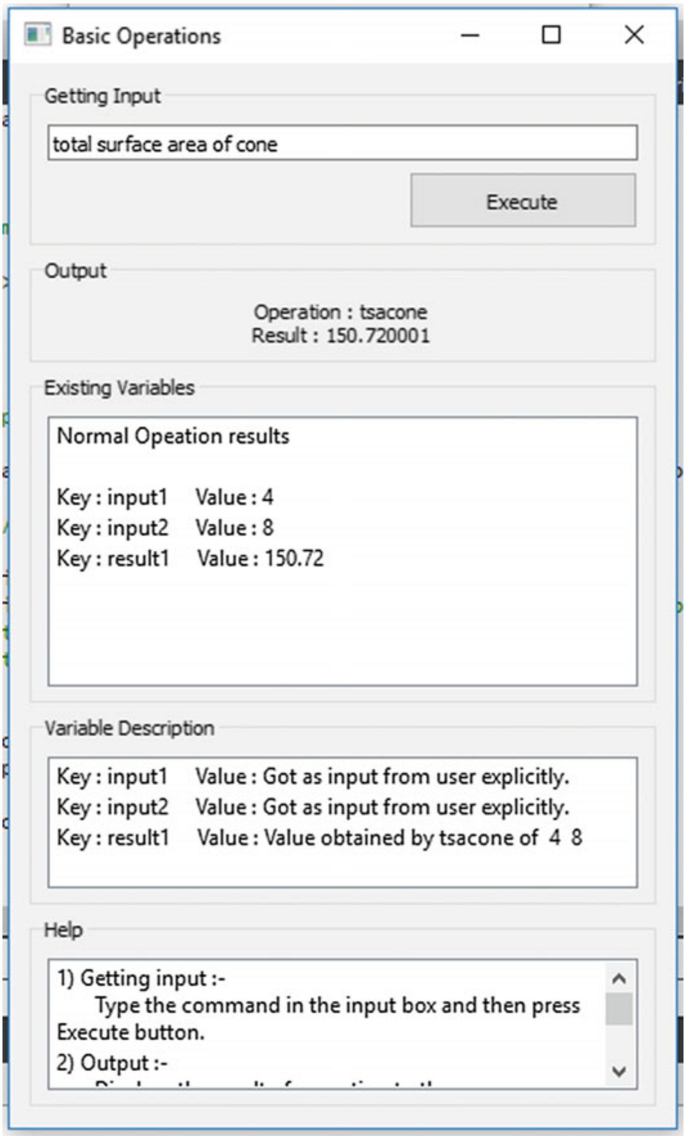


Fig. 5 Total surface area performed without values specified

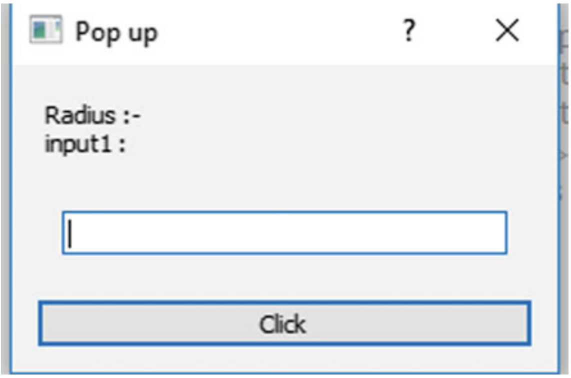


Fig. 6 Prompt box to enter value

Table 1 Characteristics and limitations of similar tools

S. No	Tool name	Input	Output	Limitations
1	Flowgortihm	Flowchart	Complete code	Altering a flowchart requires redrawing. Creating a flowchart for complex programs is difficult
2	Raptor	Flowchart with structured commands	Compiled output	Input should follow a predefined structure inside the building blocks of flowchart
3	Anchor pseudocode compiler	Code without curly braces and semicolon	Code with curly braces and semicolon	Requires programming knowledge. User has to specify complete program which abides syntactical rules with an exemption for semicolon and curly braces
4	Visual paradigm	UML diagram	C++ code	Knowledge about constructing UML diagrams is needed. The main focus is on the design
5	Codesmith generator	Code modified with template	Complete code	Searching for the function in the template is time-consuming. Function should be defined by the user

not require programming knowledge which is contradictory to Anchor Pseudocode Compiler. Visual Paradigm's primary focus is on design. On the other hand, the proposed tool focuses on completing the task rather than concentrating on design. In Codesmith Generator, function name will be present but it will not be defined in the template. However, in ONESTOP, user need not define the function as it is already defined in the tool. So, specifying the name of the operation to be done is sufficient.

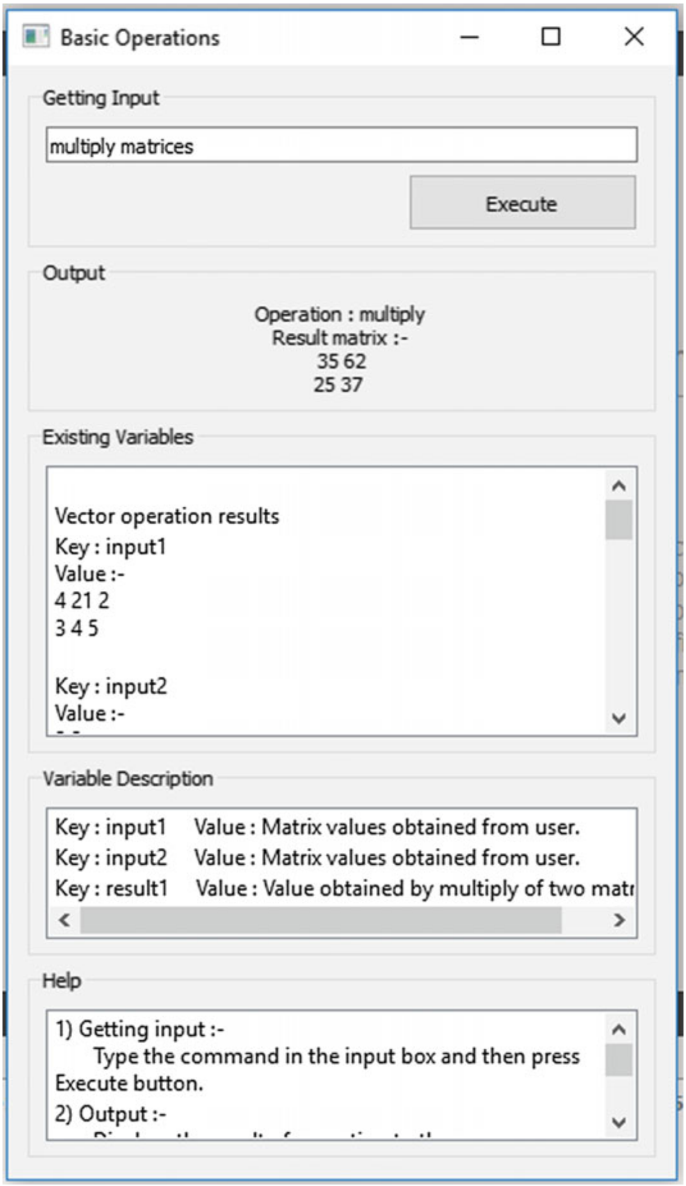


Fig. 7 Multiplication of matrices

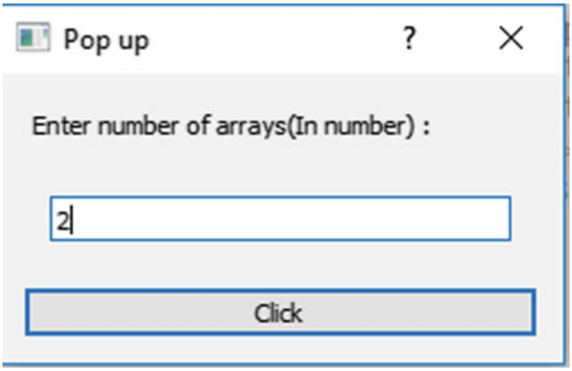


Fig. 8 Prompt box to specify number of matrices

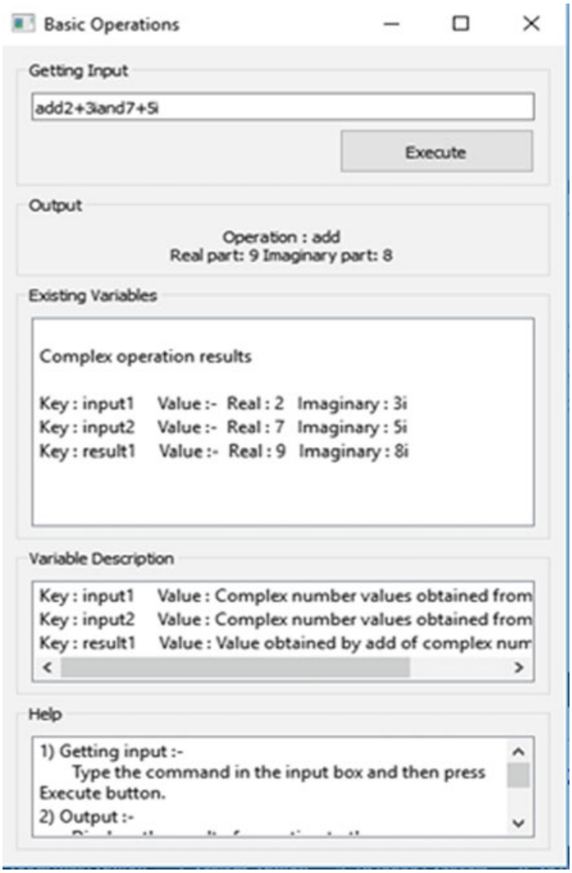


Fig. 9 Prompt box to enter matrix values

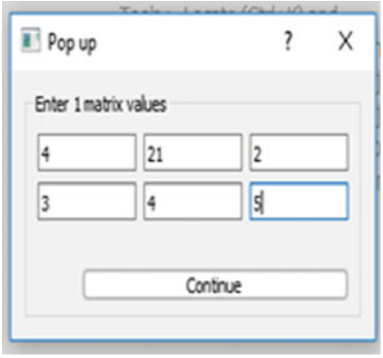


Fig. 10 Addition of complex numbers

Table 2 Mean values of scores from the survey

Features	Mean value
Usability	8.13
Lack of learning	8.45
Performance	7.72
Enhancement	6.9
Result interpretation	7.54
Tool design	7.86
Scope	6.27

5.3 *User Experiment*

A survey was conducted among few people to rate various features of the tool. These people were of three categories: novice, intermediate and experienced programmers. They were asked to rate the features of the tool on a scale of 1–10. Table 2 shows the mean value of every feature of the tool obtained from the survey conducted among three categories of people. Usability denotes how easily the tool can be handled for performing various operations. Lack of learning indicates the amount of prerequisite knowledge required for using the tool. It also implies the efforts needed to get familiarized with the tool. Performance shows how fast the processing and execution of desired operation take place to generate the output. Enhancement signifies how much the tool can be developed further. It includes adding new features. Result interpretation is how far the user is able to understand and interpret the result. Tool design indicates how attractive the front end of the tool is. Scope defines the types and number of operations that can be currently performed by the tool.

From the results in Table 2, few conclusions can be made. Highest mean value (8.45) was for lack of learning. Usability has got the second highest mean value of 8.13 because tool is helpful for novice, intermediate and experienced programmers. Scope received the least mean value since the tool is developed for performing limited operations only. Various other types of operations can be added to the tool for further development.

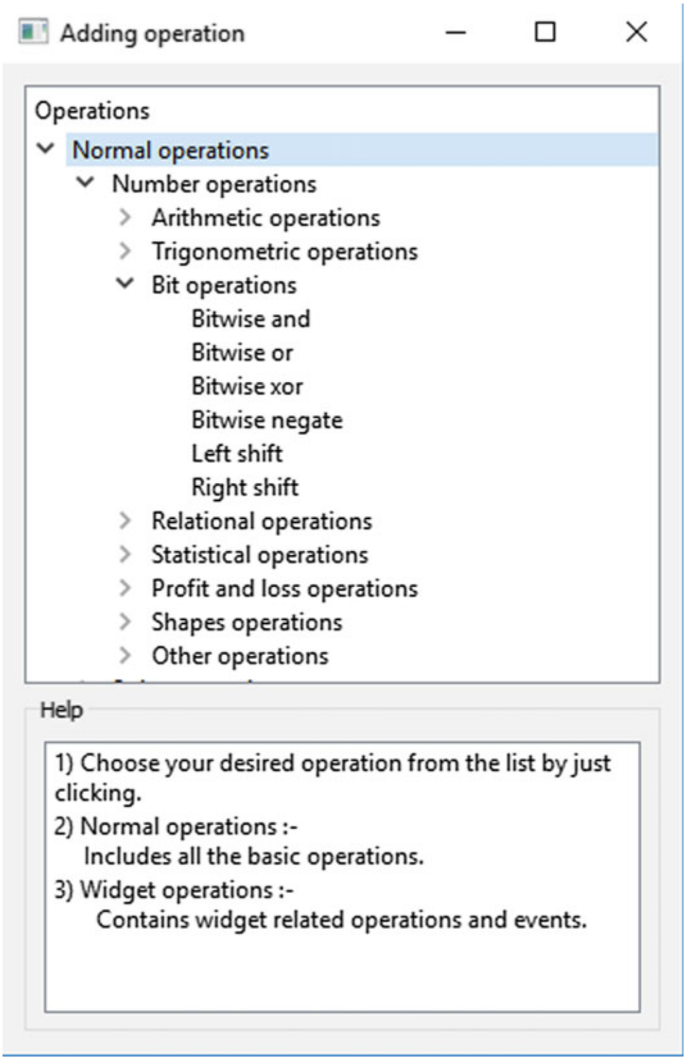


Fig. 11 Operations list

Two features which have the highest mean values are considered for further conclusion [23]. Therefore, lack of learning and usability are taken into consideration. Naive Bayes classifier is used to identify the probability of each category of people who have rated two features above 6. Posterior probability $P(c/x)$ is found by

$$P(c/x) = P(x/c) * P(c)/P(x)$$

where c is the class and x is the predictor. $P(c/x)$ is posterior probability, $P(x/c)$ is likelihood probability, $P(c)$ is class prior probability and $P(x)$ is predictor prior probability.

Let novice, intermediate and experienced programmers be the three classes. Class prior probability for each class obtained from the survey is given as

P (novice) = 0.4.
P (intermediate) = 0.31.
P (experienced) = 0.27.

Likelihood probability is calculated to find the probability of having a score greater than 6 for lack of learning and usability for a particular class. Likelihood probability as calculated as follows:

P (lack of learning > 6 and usability > 6 /novice) = 0.28.
P (lack of learning > 6 and usability > 6/intermediate) = 0.19.
P (lack of learning > 6 and usability > 6 /experienced) = 0.15.
Predictor prior probability is found out as
P (lack of learning > 6 and usability > 6) = 0.62.
Posterior probabilities are calculated as follows:
P (novice/lack of learning > 6 and usability > 6) = 0.45.
P (intermediate/lack of learning > 6 and usability > 6) = 0.30.
P (experienced/lack of learning > 6 and usability > 6) = 0.24.

From the above posterior probabilities, conclusions can be made that novice programmers have liked the lack of learning and usability features of the tool the most and have rated them mostly above 6 on a scale of 10. Intermediate programmers have the second highest probability of liking these two features. Experienced programmers have the least probability of rating these two features above 6 compared to the other two categories of people as they have sound knowledge in programming.

6 Future Work

ONESTOP can further be developed to perform the operations and produce the result from the commands got in many other languages since the tool developed is used to process sentences in English only. The main keyword can be in any language and the keyword identified from the input should be able to match the main keyword in order to identify the type of operation. Step-by-step processing of the input can be illustrated using animation as it will be easier for beginners to understand the working of the tool better. Graphical User Interface (GUI) is used predominantly to make a tool attractive. GUI features can be included to make designing easier. Based on the GUI requirements specified by users, ONESTOP can be enhanced to create GUI automatically. This can further be made easier for users by using drag and drop option to create GUI. Similar to Google search, already specified inputs can be stored and suggested to users while they type the sentence.

7 Conclusion

A tool which performs many operations based on the input string from user is created. The tool identifies keywords from input and performs the operation to produce the expected result. This tool is very easy to use. The user does not require specific knowledge about how to use the tool. ONESTOP was mainly developed to eradicate the fear of using or learning new programming languages and remembering their syntax. User need not learn programming languages to do various operations. A pseudocode or even a sentence can be used to perform an operation using this tool. Since the values of inputs and results are displayed in a sequential manner after performing the operation, it helps users to understand the logic of how the result was got.

References

1. Fidge C, Teague D (2009) Losing their marbles: syntax-free programming for assessing problem-solving skills. In: Conferences in research and practice in information technology series 95
2. Charntaweekhun K, Wangsiripitak S (2006) Visual programming using flowchart. In: 2006 international symposium on communications and information technologies, Bangkok, pp 1062–1065. DOI: 10.1109/ ISCIT.2006.339940
3. Sarda N, Jain A (2001) Mragyati: a system for keyword-based searching in databases. In: cs.DB/0110052
4. Little G, Miller R Syntax-free programming. URL http://up.csail.mit.edu/projects/keyword-commands/index.html
5. Smith T, Liu J, Roman E (2009) Final project: classification and keyword extraction of Anne Hunter 'jobslist' Emails
6. Fudaba H et al (2015) Pseudogen: a tool to automatically generate pseudocode from source code. In: 2015 30th IEEE/ACM international conference on automated software engineering (ASE), Lincoln, NE, pp 824–829
7. Tahir M, Ahmed S (2015) Tree-combined Trie: a compressed data structure for fast IP address lookup. Int J Adv Comput Sci Appl 6. https://doi.org/10.14569/ijacsa.2015.061223
8. Senthil Kumar T, Ohhm Prakash KI (2015) A queueing model for e-Learning system. In: Suresh L, Dash S, Panigrahi B (eds) Artificial intelligence and evolutionary algorithms in engineering systems. Advances in intelligent systems and computing, vol 325. Springer, New Delhi
9. Sankar A, Bharathi PD, Midhun M, Vijay K, Senthil Kumar T (2016) A conjectural study on machine learning algorithms. In: Suresh L, Panigrahi B (eds) Proceedings of the international conference on soft computing systems. Advances in intelligent systems and computing, vol 397. Springer, New Delhi

Performance Evaluation of DCT, DWT and SPIHT Techniques for Medical Image Compression

M. Laxmi Prasanna Rani, G. Sasibhushana Rao and B. Prabhakara Rao

Abstract Medical imaging is, visible illustration of internal of the human body in digital form. There exists a need for compression of these medical images for storage and transmission purposes with high image quality for error free diagnosis of diseases. It is necessary to develop new techniques for compression of images, resulting into reduction in cost of data storage and transmission. This paper proposes a progressive and DWT-based Set Partition in Hierarchical Tree (SPIHT) algorithm for achieving better image compression while maintaining the nature of restored image. The performance of SPIHT algorithm is evaluated and compared with simple DCT and DWT using the parameters like Peak Signal to Noise Ratio (PSNR), Mean Squared Error (MSE) and compression ratio (CR). From the results it is concluded that SPIHT algorithm produces better PSNR and MSE values compared to DCT and DWT with high image quality.

1 Introduction

Image compression is one of the important applications of digital image processing for better data transmission and storage. Generally this image data require abundant memory space for storage and higher bandwidths for transmission. To reduce the data required for digital image representation, the method of compression of image is necessary. Different applications of communications like Television, satellite and error free diagnosis of diseases in medical applications, storage and communication

M. Laxmi Prasanna Rani (✉)
Department of ECE, MVGR College of Engineering, Vizianagaram, Andhra Pradesh, India
e-mail: prassugowtham@gmail.com

G. Sasibhushana Rao
Department of ECE, AUCE(A), Andhra University, Visakhapatnam, Andhra Pradesh, India
e-mail: sasigps@gmail.com

B. Prabhakara Rao
Department of ECE, JNTUK, Kakinada, Andhra Pradesh, India
e-mail: drbprjntuk@gmail.com

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_147

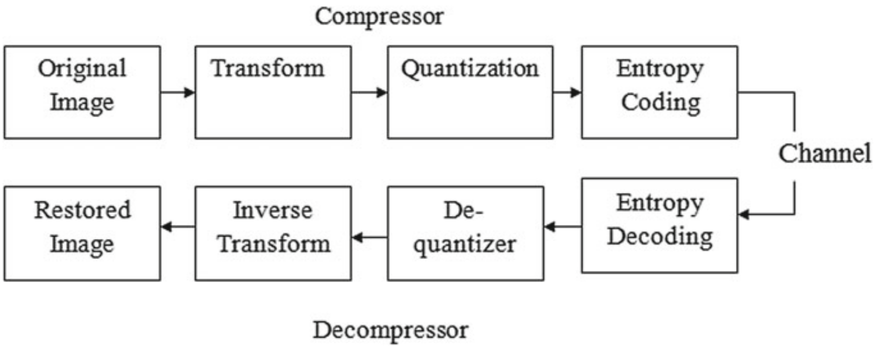


Fig. 1 Block diagram of Lossy compression

will not be feasible without compression of data [1]. The objective of an image compression technique is to reduce the repeated and irrelevant data in images such that small numbers of bits are used to represent the image in the vicinity of an "acceptable" quality of vision for the reconstructed image. By reducing the redundancy and irrelevance, the data will be stored and transmitted in effective form which is known as compression. The compression methods are also suitable for the devices have less on board memory and low battery life.

This paper recommends an image compression technique for the transmission of data in the applications of communications and wireless network. This should be maintained by balancing available bandwidth and perceived good quality of received image with minimum delay of transmission. To get better compression, the factors like algorithm complexity, compression time, computational resources and cost to be considered. These factors are essential and to be considered design or modify a compression method to make it fast and accurate to get good quality images.

The image compression techniques are of two type's i.e. lossy or lossless compression. Lossy compression techniques provide considerable decrease in data rate [2] with loss of accuracy for natural images such as photographs. Lossless compression techniques are suitable for medical images by lowering the no. of bits of the data without reducing the eminence of the image. Various image compression techniques such as transform coding and adaptive versions of techniques have developed. The block diagrams of lossy and lossless compressions are shown in Figs. 1 and 2 respectively [2].

DCT is a common compression technique for JPEG images based on block processing methods [3]. DCT represents with sum of cosine functions of an image varies at different frequencies and magnitudes. This transform has particular property of energy compaction. Image compression using the DCT is simple and get better compression ratio (CR).

The Discrete Wavelet Transform (DWT) is an efficient transform tool for compression of images due to its capability of displaying the images at different resolution also gives better compression. This wavelet transform represents with a sum

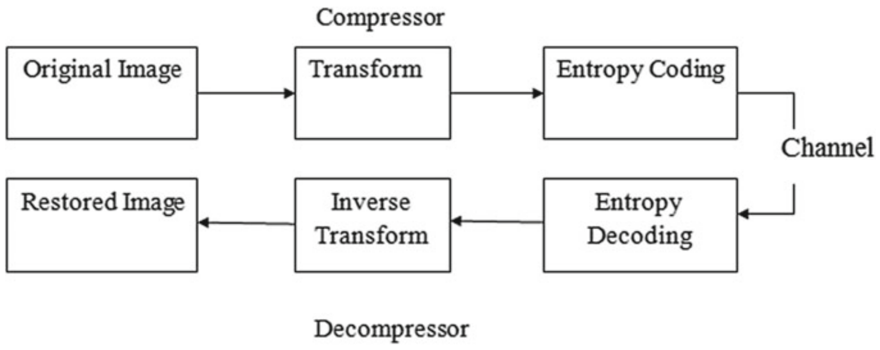


Fig. 2 Block diagram of loss less compression

of wavelet functions at different resolution of an image due to the property of Multi Resolution Analysis. The basis function of this transform is of any function which satisfies the requirements of multi resolution analysis [2].

Set Partition in Hierarchical Tree (SPIHT) is one of the best compression algorithms which reconstruct the original image with low loss and provides good image quality. This algorithm provides excellent PSNR and improved CR and low encoding complexity than DWT and DCT [3].

This paper is organized such that introduction to discrete cosine transform, discrete wavelet transformation and SPIHT algorithm of Image is described in Sects. 1, 2 describes SPIHT algorithm and Sect. 3 illustrates the compression using SPIHT algorithm. Section 4 discusses the results and conclusions are presented in Sect. 5.

2 SPIHT Algorithm

SPIHT algorithm is award-winning compression technique has received worldwide applause and consideration since 1995. Many researchers have tested and used this algorithm which becomes the standard and advanced algorithm for compression of images. This technique is very fast and is the best among other compression techniques. It is a progressive hierarchical structure of image transmission and gives better Image eminence with high PSNR. Using wavelet transform the image is decomposed into various sub-bands as the initial step of this algorithm. The decomposition is a repetitive process until it reaches the final value. For each decomposition, there are four sub-bands i.e. approximation coefficients of low-frequency and horizontal, vertical, diagonal coefficients of high-frequency. The coefficients of different sub bands related to the same spatial position display self-similarity characteristics in the pyramid structure [4]. SPIHT establishes spatial orientation trees by defining parent children relationships between these self- similar sub bands [5]. The spatial orientation tree is an example of parent–children relationship is shown in Fig. 3.

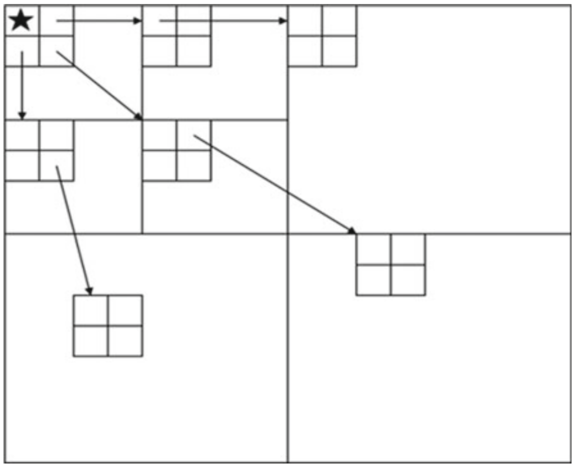


Fig. 3 Examples of parent-offspring dependencies in the spatial-orientation tree

2.1 Steps in SPIHT Algorithm

There are three steps of sorting, refinement and quantization in SPIHT algorithm. Here, the image is encoded and differentiated into significant and insignificant pixels using three lists such as LIS (Least Insignificant Set), LIP (Least Insignificant Pixel), and LSP (Least Significant Pixels). The maximum number of bits is $n_{\max}$ depends on the largest pixel value in the spatial orientation tree and it can be represented by the Eq. (1).

$$n_{\max} = \left[\log_2\left(\max\left(\left|C_{i,j}\right|\right)\right)\right] \quad (1)$$

where $C_{i,j}$ represents the largest pixel value in the spatial tree.

The coefficients in the Least Insignificant Pixels, Least Insignificant Set and Least Significant Pixels are separated using the threshold value Th. The threshold value can be expressed in Eq. (2) given below.

$$\text{Threshold Value, Th} = \begin{cases} 1, & \max\left(\left|C_{i,j}\right|\right) \geq 2^n \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

The steps in SPIHT algorithm are explained in the following steps using three types of passes such as sorting, refinement and quantization pass [6].

Step 1 In the sorting pass, the pixels or coefficients in the list of Insignificant Pixel (LIP) are examined to determine whether the pixels or coefficients are significant or insignificant depending on threshold value Th. If the magnitude of coefficients more than the value of threshold are found to be significant, then ‘1’ is the output and the other bit is taken as the sign of the pixel coefficient (‘1’ for positive and ‘0’ for negative). Then the significant pixel coefficient is shifted to significant pixels list. If any coefficient is less than threshold value, it becomes insignificant and that bit is taken as ‘0’.

Step 2 In this step, the overall wavelet coefficients in List of Insignificant Set are to be processed. The set of all offspring’s of coefficients in the list of LIS are correlated with the threshold value to make a decision whether they are important or not. If the coefficients are greater than a certain threshold value then consider the coefficients as significant and move into least significant pixels; otherwise the coefficients moved into LIP as insignificant [7].

If the coefficient is insignificant, then the spatial orientation tree was a zero-tree, which is rooted by the current entry. Hence, the bit ‘0’ is output and there is no further processing of these coefficients.

The remaining coefficients, which are coefficient descendants set except for the immediate offsprings, are shifted to the last of List of Insignificant Set. Again thresholding test is performed on all the descendants of its direct offspring’s of coefficients. Once the significance test is done, the spatial tree again split into four sub-trees which are rooted by the coefficients of direct descendants and these direct offspring’s are added in the end of LIS.

Step 3 In the refinement pass, all the refinement bits (*n*th bit) in Least Significant Pixel at current threshold are taken as output. For the next round of this algorithm, the current threshold becomes halved. This process is continuous when all the coefficients in LIS are completed and the quantization can be done to get compression.

The implementation steps of SPIHT algorithm is explained in Fig. 4.

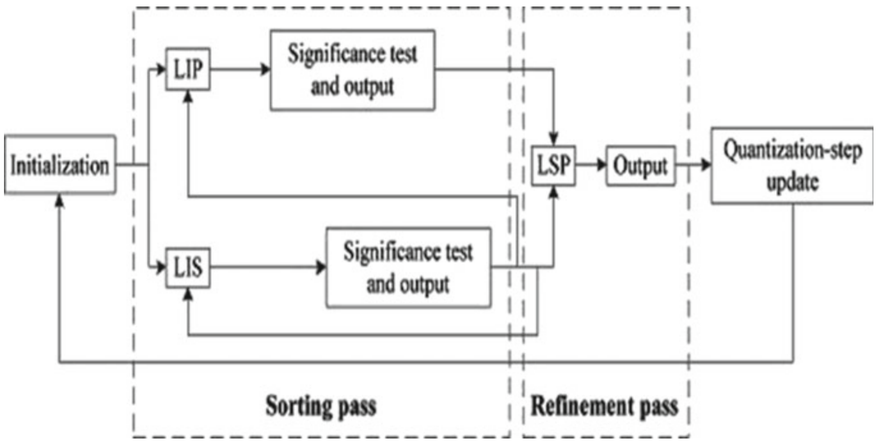


Fig. 4 SPIHT algorithm implementation steps

3 Image Compression Using SPIHT Algorithm

Compressing Images with SPIHT technique can be explained below. Initially, image is decomposed into approximation coefficients (LL) of low frequency and horizontal (LH), vertical (HL), diagonal (HH) coefficients of high frequency, by means of wavelet transform. The approximation coefficients further divided into different components depending on level of decomposition and the level equals the value of n.

The image data is encoded in SPIHT algorithm [8] after wavelet decomposition in the steps of sorting, refinement and quantization using three lists such as LIP (Least Insignificant Pixels), LIS (Least Insignificant Set) and LSP (Least Significant Pixel). When the magnitudes of coefficients less than threshold value are in the list of LIP, overall wavelet coefficients defined in the tree structure less than threshold are in LIS and finally LSP contains the set of pixel coefficients having magnitude higher than threshold are the important pixels [9].

In the sorting process, all the coefficients in the list of LIP are verified if they are significant or insignificant depending on threshold and the descendants of the coefficients in remaining lists are also tested using threshold value in the Eq. (2). The coefficients greater than thresholds are inserted into LSP and remaining coefficients are moved to LIP [5].

In the refinement process, the nth MSB of the coefficient in the LSP is taken as the final output. Next the value of n is decreased by '1'. Again sorting and refinement processes are applied to coefficients descendants until the value of n reaches '0' or the list coefficients in the LIS are completed. The quantization process is applied to the entire coefficients in list of significant pixels.

Once the encoding process is completed, the compressed image is applied to SPIHT decoding process. The true image is reconstructed with better image quality using Inverse Discrete Wavelet Transform. This complete process is explained in Fig. 5.

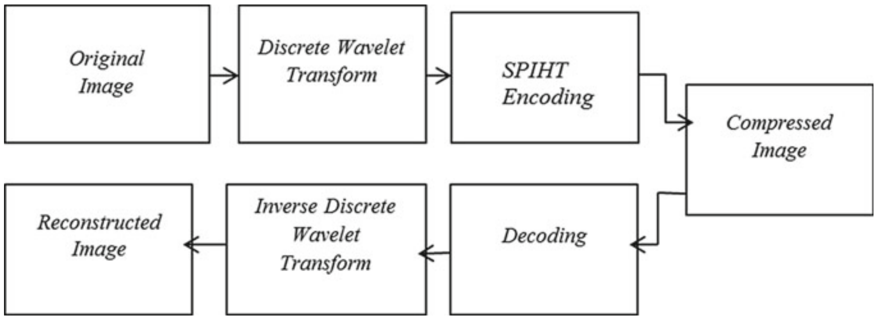


Fig. 5 Image compression using SPIHT algorithm

3.1 Performance Metrics

The performance of compressing the images using SPIHT can be obtained by calculating PSNR (Peak Signal to Noise Ratio), MSE (Mean Squared Error) and the compressed image size. PSNR is used to determine the excellence of the image. High PSNR value resembles that the reconstructed image quality is high and low PSNR value gives that reconstructed image quality is low. MSE is another parameter that measures the excellence of the image by estimating the difference between original and reconstructed images [10].

$$\mathrm{MSE} = \frac{1}{MN}\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}\left((p(x,y)-s(x,y))^2\right) \tag{3}$$

where $p(x, y)$, $s(x, y)$ are original and reconstructed images with size of *MXN*.

$$\mathrm{PSNR} = 10\log_{10}\left(\frac{I_{\max}^2}{\mathrm{MSE}}\right) \tag{4}$$

where $I_{\max}$ is the maximum possible pixel values of the image.

4 Results and Discussion

This paper presents a progressive compression algorithm i.e. SPIHT algorithm applied to the medical images like MRI of brain, X-ray images and a test image of lena. This algorithm is implemented in Matlab and the results are compared with present strategies of DCT and DWT. Performance of SPIHT algorithm is evaluated using the parameters of Peak Signal to Noise Ratio (PSNR), Mean Squared Error (MSE) and sizes of input and compressed image sizes.

Table 1 gives the sizes of original and compressed images using DCT, DWT and SPIHT. Table 2 shows the PSNR value obtained using MSE calculated for medical images using DCT, DWT and SPIHT algorithm respectively. Figure 6 shows the original and compressed images of the existing techniques of DCT and DWT and progressive hierarchical structure of SPIHT algorithm. From the results, it is observed that better compression obtained by DCT over DWT and better reconstruction is obtained by SPIHT algorithm with more PSNR and less MSE.

5 Conclusion

This paper progressed with SPIHT algorithm to get better image compression with progressive hierarchical structure. The values of PSNR and MSE obtained from

Table 1 Comparisons of original and compressed images sizes

Reference image	Size original image (KB)	Compressed image size(KB)		
		DCT	DWT	SPIHT
Test image (lena)	138.056	1.329	3.964	28.686
MRI (Brain1)	19.699	15.453	17.557	18.714
MRI (Brain2)	74.889	10.116	16.871	27.593
X-ray	51.462	5.077	5.016	28.342

Table 2 Comparison of MSE and PSNR values of DCT, DWT and SPIHT

Reference image	DCT		DWT		DWT Based SPIHT	
	MSE	PSNR (dB)	MSE	PSNR (dB)	MSE	PSNR (dB)
Test image (Lena)	34.2357	24.1008	29.5102	34.3105	4.2156	41.82
MRI (Brain1)	10.5186	37.9112	4.3304	43.9405	0.7534	49.36
MRI (Brain2)	1.3593	46.7975	0.2599	53.9832	0.2074	54.96
X-ray	9.2746	38.4579	0.4411	51.6851	0.0411	55.6851

Fig. 6 (i) Original. (ii), (iii), (iv) Compressed images of DCT, DWT and SPIHT of lena

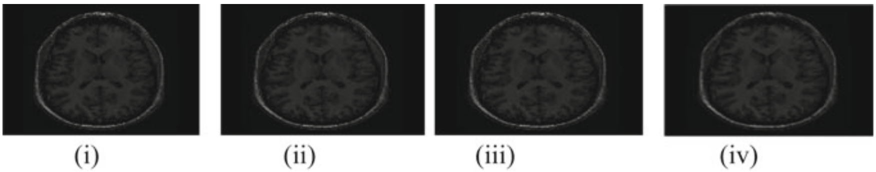


Fig. 7 (i) Original. (ii), (iii), (iv) Compressed images of DCT, DWT and SPIHT of MRI of brain1

SPIHT, is compared with the results of DCT and DWT. From these results SPIHT algorithm exhibits good image compression and also reconstructs the original image without reducing the image quality. From the testing outcomes, it could be concluded that the SPIHT algorithm is successful and capable than prevailing strategies in phrases of PSNR, MSE (Figs. 7 and 8).

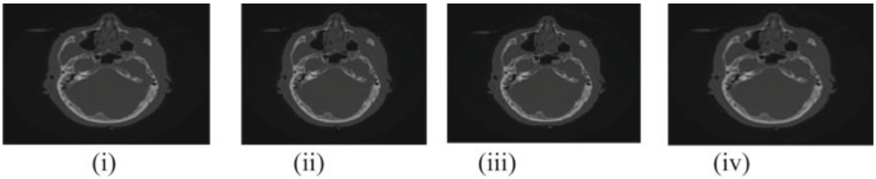


Fig. 8 (i) Original. (ii), (iii), (iv) Compressed images of DCT, DWT and SPIHT of MRI of brain2

References

1. Kulkarni, AN, Gandhe ST, Dhulekar PA, Phade GM (2015) Fractal image compression using genetic algorithm with ranking select mechanism. In: International conference on communication information & computing technology (ICCICT)
2. Gonzalez RC, Woods RE (2009) Digital image processing. Pearson Education, 2nd edn
3. Telagarapu P (2011) Image compression using DCT and wavelet transformations. Int J Sig Process Image Process Pattern Recogn 4(3)
4. Shapiro JM (1997) Embedded image coding using zero trees of wavelet coefficients, pp 108–121
5. Venkataramani DY, Banu SP (2011) An efficient hybrid image compression scheme based on correlation of pixels for storage and transmission of image. Int J Comput Appl 18(3):0975–8887
6. Rehman M, Touqir I, Batool W (2015) Selection of optimal wavelet bases for image compression using SPIHT algorithm. In: Seventh international conference on machine vision (ICMV 2014)
7. Yang SH, Chang YL, Chen HC (2001) A digital watermarking scheme based on SPIHT coding. In: 2001 IEEE international conference on multimedia & expo (ICME 2001). Tokyo, Japan
8. Said A, Pearlman WA (1996) A new, fast, and efficient image codec based on set partitioning in hierarchical trees. IEEE Trans Circ Syst Video Technol 6(3)
9. Ezhilarasi P, Nirmalkumar P (2013) Efficient image compression algorithm using modified IWT and SPIHT for CMOS image sensor. Int Rev Comput Softw
10. Gu X, Ren K, Andreas JM, Hielscher H (2009) Parametric image reconstruction using the discrete cosine transform for optical tomograph. J Biomed Opt

Structural Health Monitoring—An Integrated Approach for Vibration Analysis with Wireless Sensors to Steel Structure Using Image Processing

C. Harinath Reddy, K. M. Mini and N. Radhika

Abstract Wireless sensors for structural health monitoring (SHM) with a coordinated approach to steel structure has obtained research area in the field of structural engineering due to its low cost and a wide range of applications. These systems have been used to monitor structural behaviour, and it has the potentiality to improve structural life period and also to improve public security too. In this paper, the study is done through low-cost wireless vibration sensors that are deployed on steel frame structure for vibration analysis using microcontrollers and data systems. In this project, tests were done to interface a wireless sensor to an Arduino. And this sensor data to PC via serial and visualize the data in a software called image processing in Arduino as a signal output. This paper explains the real-time deployment of SHM using sensor networks.

1 Introduction

Detecting the structural behaviour or operation of the method of implementing a property and enactment for structures is termed to as Structural Health Monitoring (SHM). Over the last few years, Wireless Sensor Networks (WSNs) have become evident and as an economical for connecting sensor networks [1]. Here, detection

C. Harinath Reddy (✉) · K. M. Mini
Department of Civil Engineering, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Coimbatore, India
e-mail: harichinnavula@gmail.com

K. M. Mini
e-mail: k_mini@cb.amrita.edu

N. Radhika
Department of Computer Science and Engineering, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Coimbatore, India
e-mail: n_radhika@cb.amrita.edu

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_148

and characterization are interpreted as changes to geometric properties of a structure, including changes to structural functioning with properties in taking loads. Especially in conditions of loads to structures such as earthquakes or blast loading, SHM is used for model analysis testing and performance of the structure can be attained based on monitoring techniques.

Wireless sensors are used for sensing, communication and control, and provide an attractive alternative for structural mitigation. These wireless control systems have flexible installation, rapid deployment and low maintenance cost. The characterization and field test of the structure using sensors is implemented in open hardware. The system utilizes the unique features of sensors for developing low cost, compact and higher accuracy system [2]. The fundamental approach sensors monitor the seismic response or vibration analysis of building structure. These sensors have broad range of application, to monitor and control structural systems. With the designed programs and parameters, different sensors can be inputted to different software systems and to the sensing unit, for monitoring purpose [3].

The vibration analysis-based detection technique is the health monitoring parameter and is implemented using wireless sensors, with data acquisition system [4]. The term detecting technique infers a total loss of system functionality or property behaviour of structure [5, 6]. Structural property change in normal service may include stiffness of structure or earthquake vibrations [7]. Structural stiffness and strength reduction are caused due to elements of the structure when in vibrations or oscillations. Many works are already carried on the structural health monitoring of civil structures using wireless sensing technology [8–12].

In this paper, the microcontroller named Arduino UNO board is used and connected to the SW-420 vibration sensor with 5 V relay power supply unit. The working parameters are programmed in the Arduino software and compiled in the microcontroller, using the board software. With performance evaluation, results are displayed in Arduino software. In this approach, for output visualization, image processing tools with Arduino and Matlab 2013a software are shown. Since the output experimentally can be shown digitally by values in order, **image signal processing** is attained by using wireless sensors networks. It represents the set of characteristics or parameters related to the tests conducted as per the sensor deployed for monitoring.

The essential principle is to compare the frequency/time (dynamic analysis) with vibration and amplitude time simulation parameter. The results are obtained through evaluation of stiffness of the material steel frame with Matlab simulation. Both experimental and computational approaches were used for frame analysis.

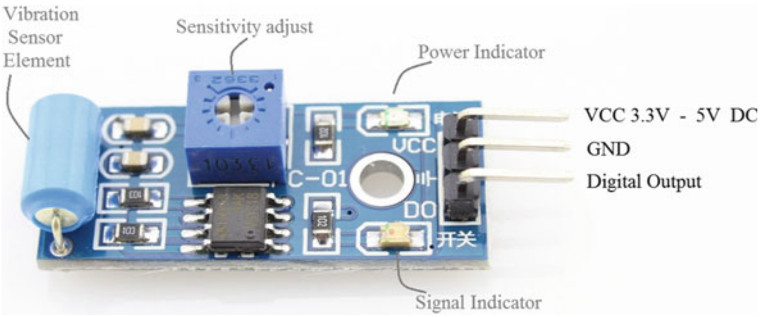


Fig. 1 SW-420 vibration sensor module

2 Materials and Parameters

Material	Sensor used	Parameters monitoring
Steel frame model-modelling on vibration motor Table. Three nos 1.5–3 mm size steel/Fe plates with 4 nos column rods of dia ϕ 10 and 12 mm	(a) Vibration SW-420 sensor	Frequency/time (dynamic analysis) (i) vibration speed versus frequency (ω) Matlab: (ii) amplitude versus time (iii) magnitude versus frequency (kHz) (iv) intensity or vibration response using FFT spectrum

2.1 Wireless Sensor System

Description:

SW-420 Vibration Sensor Arduino (Fig. 1):
The vibration sensor SW-420 is composed of jumper pin board with adjustable comparative system onboard potentiometer. This is used for sensitivity conditioning and signal indication light emitting diode (LED).

When vibration or an external force applied on the sensor module, it senses on the vibration motion, and then the output of this module gives the signal.

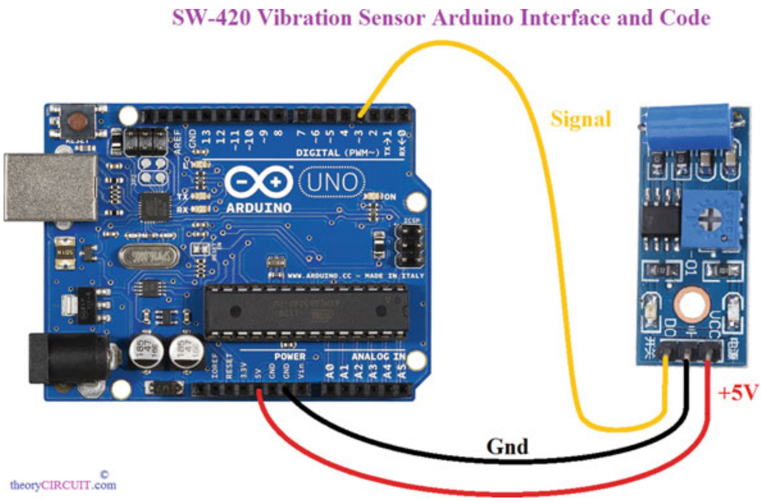


Fig. 2 Arduino circuit connection with SW-420 interface

Arduino Hookup with SW-420 (Fig. 2)
Sensor board Vcc (voltage at the common collector) pin is connected to 5 V pin; connect Gnd (ground) pin to Gnd pin of Arduino board; and connect DO (digital output) signal pin of sensor board to Arduino digital pin A3. By doing the calibration, the sensitivity threshold is adjusted.

Using USB cable connect, the entire interface of Arduino mote to the data acquisition (DAQ) system is used to visualize signals.

Experimental Investigation: Steel Frame Model test Equipment Portal frame is with three-storey (G + 3) and 1 m ht, Arduino UNO circuit board, SW-420 vibration sensor, vibration motor table and data acquisition system.

2.2 Parameter for Monitoring Vibration of Portal Frame Using Wireless Sensors

Frequency/time (dynamic analysis), vibration and amplitude, time simulation technique, 3-storey steel frame are used in a model for experimental study of structural health monitoring. Basic aspect is considering dynamic analysis which is different from static analysis. In this investigation of experiment tests, dynamic loads are applied as a function of time or frequency, and this condition of varying load application induces time- or frequency-varying response, i.e. accelerations, displacements, etc. These test characteristics make dynamic analysis more complicated and realistic than static analysis. So here in the experimental test, load application is created due to vibration motor and induces a dynamic response to steel structure. Installed

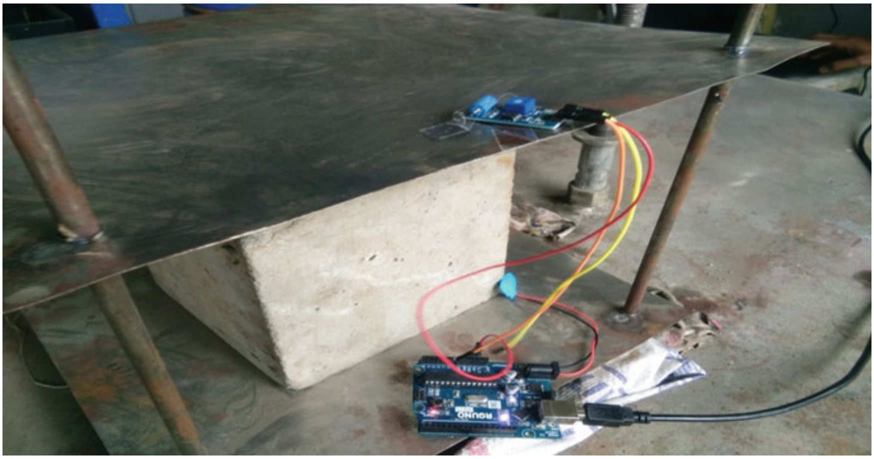


Fig. 3 Sensor position at first storey and restrained base plate in vertical direction (Y-direction) (Mechanism by concrete block)

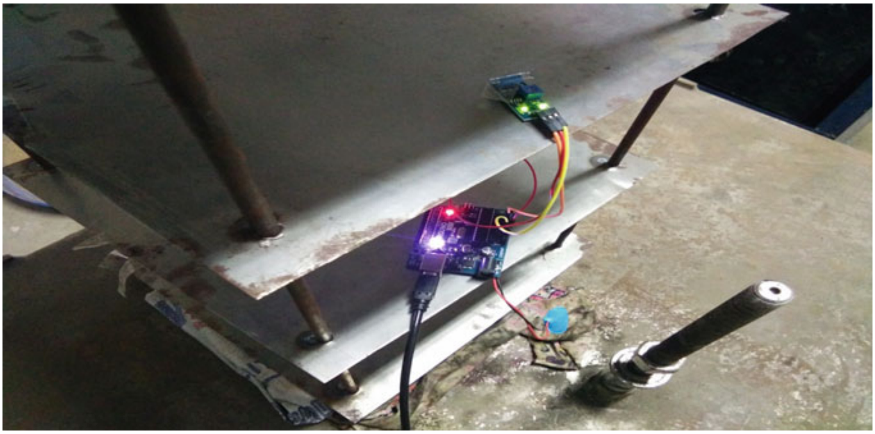


Fig. 4 Sensing interface and communication between the vibration module and Arduino with DAQ

wireless vibration sensor on three-storey frame, first storey, second storey and third storey as points (S1, S2, S3, S4…), are marked and used for testing by placing it on the vibration table (Fig. 3).

To perform and analyse the test using wireless sensing unit for the structure, active mechanism and control are implemented. These sensors kept on each floor and connected with data acquisition systems to transmit the response to the receiver (Fig. 4).

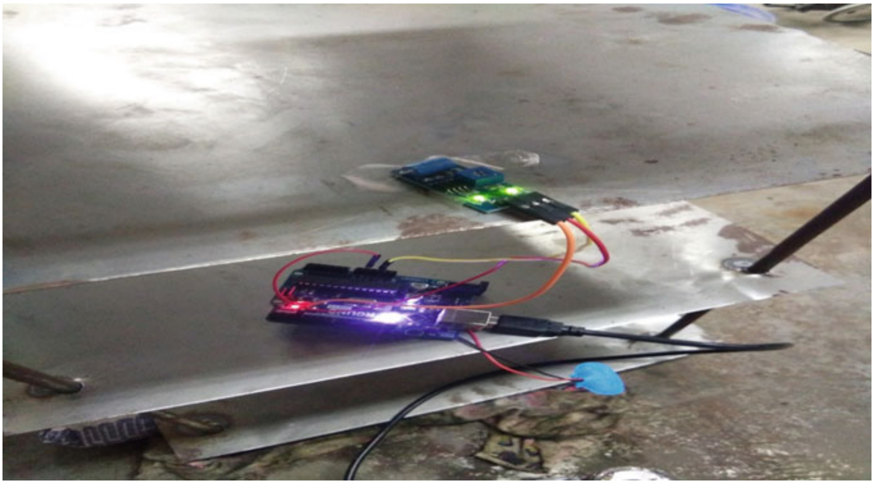


Fig. 5 Capturing of the value response of the steel structure

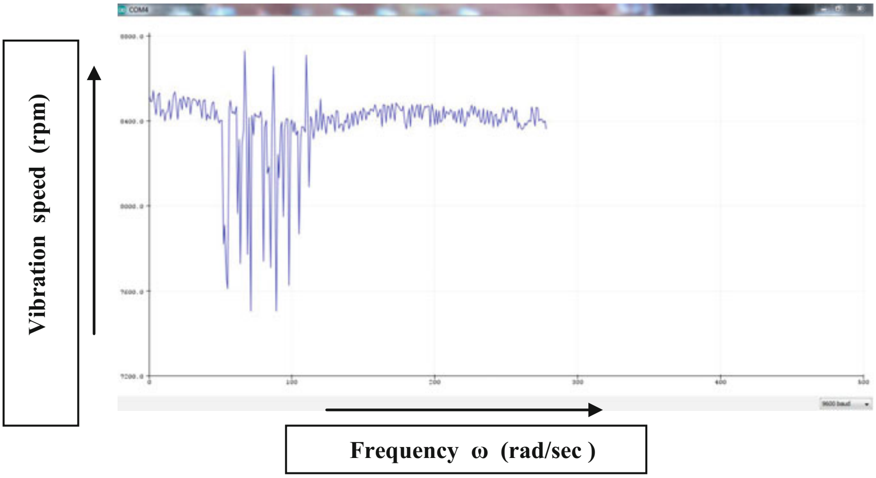


Fig. 6 Response vibration on *Y*-axis versus frequency (ω) on *X*-axis corresponding to position S1

The vibration sensor is deployed with initial arrangements and connections. The pins are connected to corresponding outputs and checked with the programmed parameters and compiled in it (Fig. 5).

Once the connections are done, sensing interface and communication are overall checked without any errors.

In Fig. 6, response signal from vibration sensor with vibration speed and frequency is attained with Arduino data acquisition system at corresponding position S1.

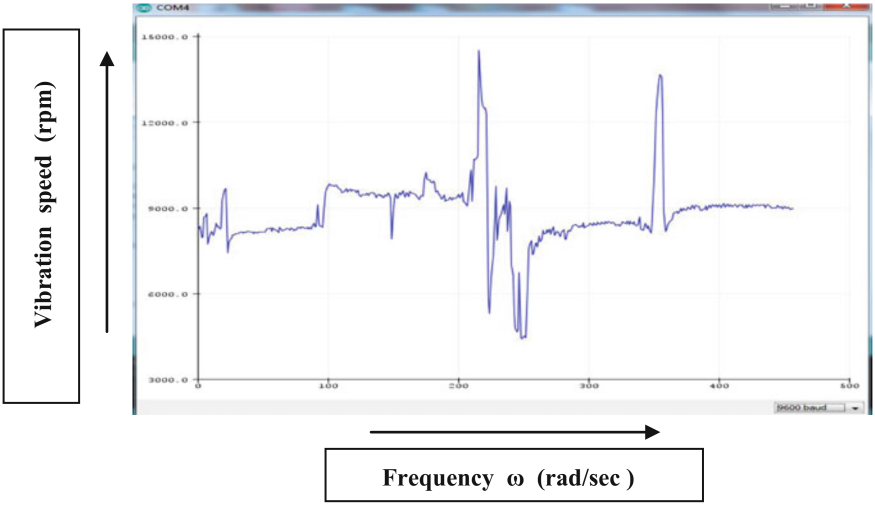


Fig. 7 Response vibration on Y-axis versus frequency (ω) on X-axis corresponding to position S2

Initial first phase of position is taken on first storey of steel frame as "S1". And the vibration motor is switched ON with certain initial speed of vibration.

At position S1, the frequency attained is captured in Arduino software from response due to sensor.

In Fig. 7, response signal from vibration sensor with vibration speed and frequency is attained with Arduino data acquisition system at corresponding position S2.

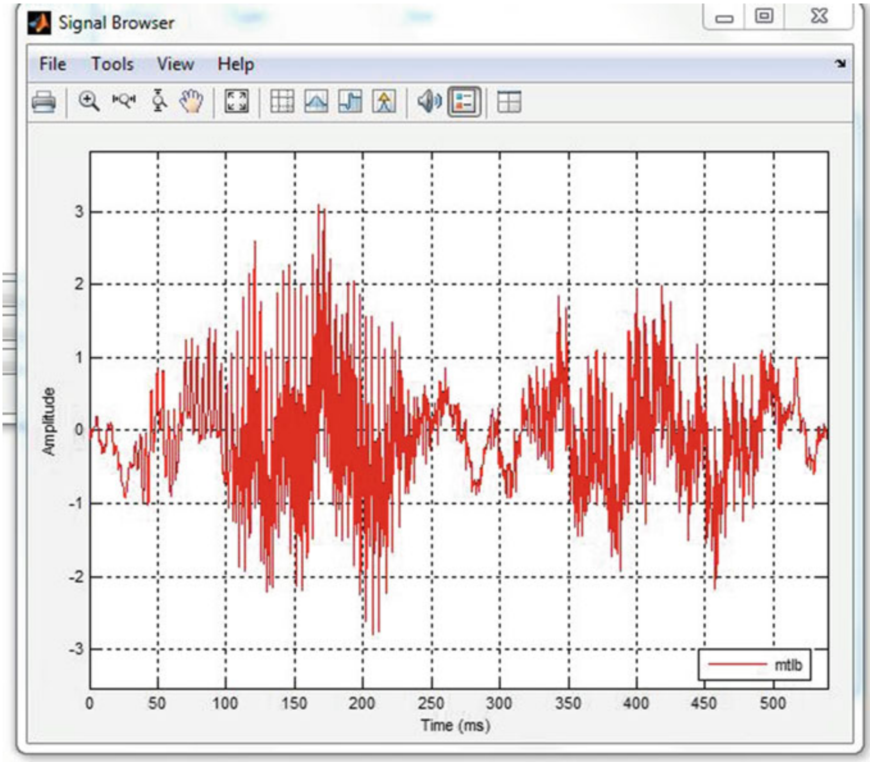
For the vibration speed of motor, at S2, considering second-phase position on second storey of steel frame, the vibration is transferred to the storey with frequency attained peak value captured in Arduino acquisition system from response due to sensor.

From this, peak value of frequency with respect to vibration speed of frame with sensors is obtained on the DAQ with Arduino and vibration sensor.

Data Obtained from Wireless Sensors to DAQ (Data Acquisition System)
Comparison between peak values are obtained from the wireless sensors measured and from those simulations using the Matlab Simulink using the measured data (Fig. 8).

From the obtained experimental test, vibration analysis for steel frame and its corresponding Matlab simulation is done by Arduino and sensor. Amplitude is taken as the parameter for initial representation of frequencies with vibration against corresponding time (Fig. 9).

For every vibration, a certain degree of magnitude is taken from an extent the steel frame attained versus corresponding frequency of motor simulation runs under which by the frame with vibration in magnitude dB (decibel) or noise (or) sound is obtained (Fig. 10).



AMPLITUDE VS TIME (ms)

Fig. 8 The Matlab Simulink with corresponding amplitude vibration of the structure with time

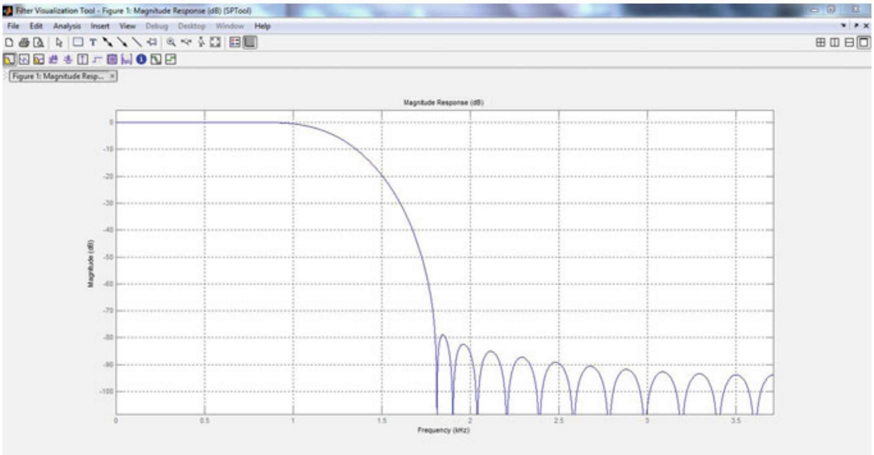
From this, FFT spectrum graph obtained from Matlab Simulink for peak values is attained corresponding to sensor values.

From peak values obtained from experimental test using vibration sensor to steel frame, the FFT spectrum estimate is compared with corresponding frequency and magnitude of vibration.

Frame Analysis—Computational Approach

From fundamental equation $M\ddot{U} + C\dot{U} + KU = 0$ and $M\ddot{U} + C\dot{U} + KU = P(t)$ or $X(t)$, 'M' is the mass of steel frame model, 'C' is the damping condition (if damper is assumed) and 'K' is the stiffness of the model.

Both real and imaginary parts are solved and are substituted in the above equation. Mathematical analysis is done using Matlab program:



Magnitude (dB) vs frequency (kHZ)

Fig. 9 Maximum attained peak value from Matlab Simulink corresponding to wireless sensor frequency value

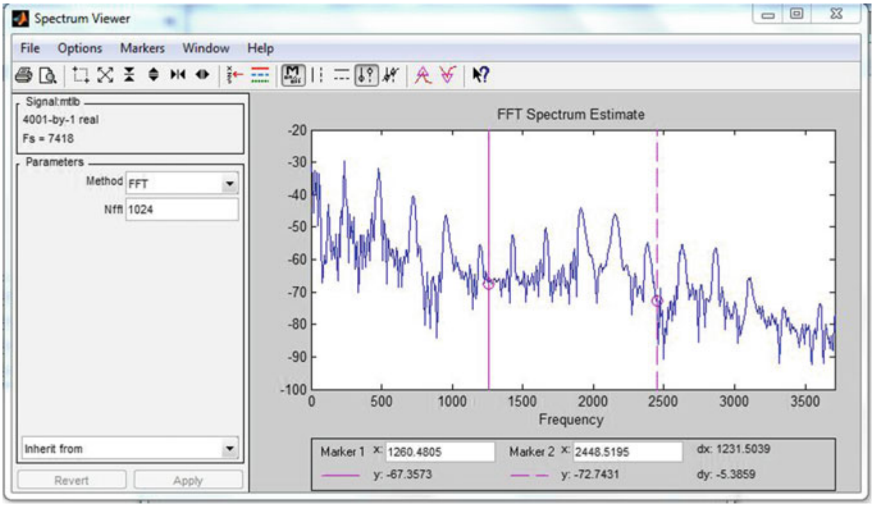
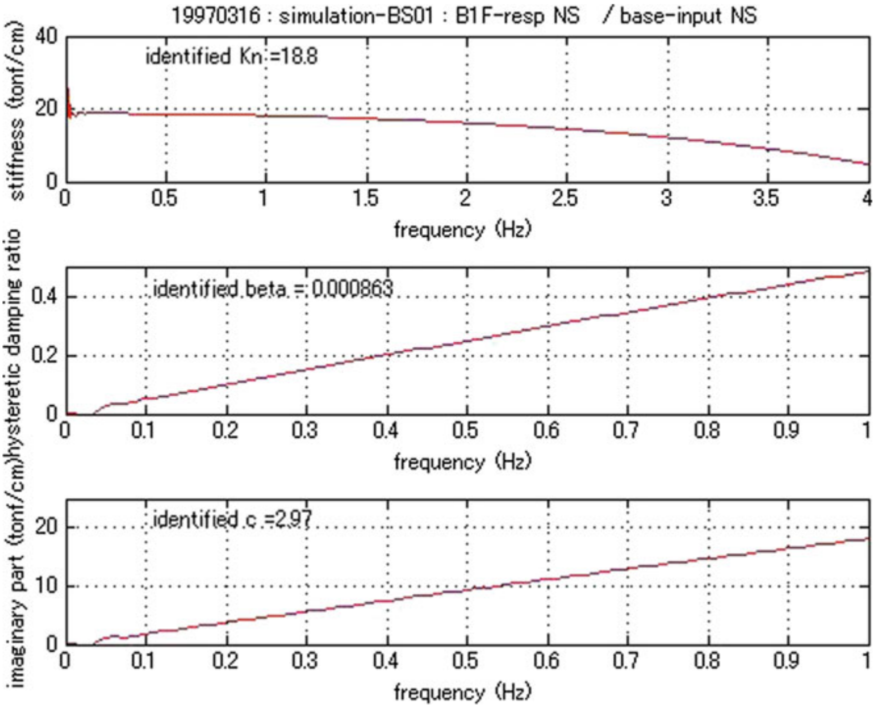
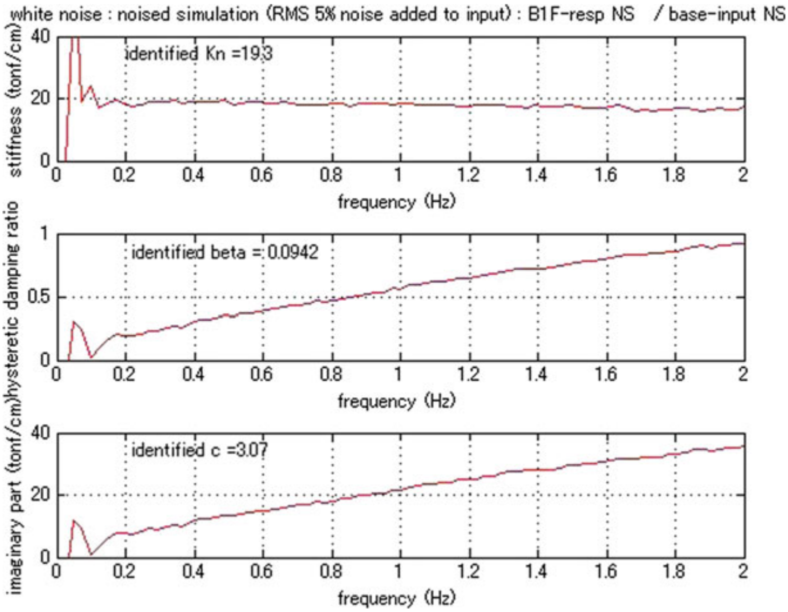


Fig. 10 Intensity or vibration response versus frequency (FFT spectrum)



From above analysis, Kn is stiffness of structure attained by simulation analysis, 'β' (beta) is the ratio of excitation frequency to natural frequency and 'C' is the damping coefficient attained from analysis.

For the steel frame model, analysis is done in Matlab using the above parameters, and stiffness (Kn) value from experimental test is compared and analysed theoretically and mathematically using dynamic analysis as obtained.



From above analysis, Kn is stiffness of structure attained by simulation analysis, 'β' (beta) is the ratio of excitation frequency to natural frequency, 'C' is the damping coefficient attained from analysis by 'NOISE' simulation approach using root mean square (RMS) method with 5% as input added (Fig. 11).

For this frame analysis, the stiffness of the steel structure (K) due to vibration analysis parameter is computationally predicted with autoregressive eXogenous (ARX) with two conditions:

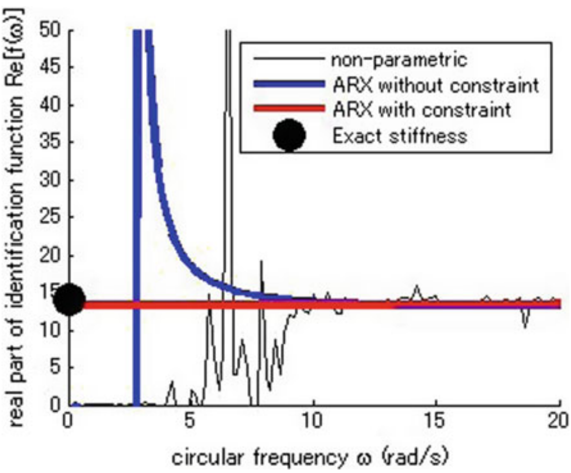


Fig. 11 Multi-loop model attained by image processing using MATLAB by auto regressive plot (ARX)

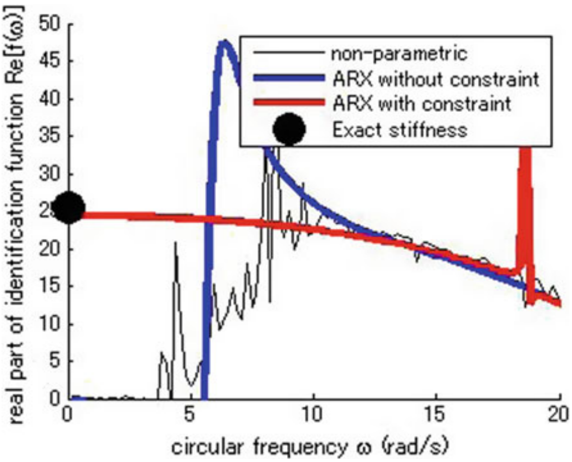


Fig. 12 Based on autoregressive—partial least squares (ARX) attained multi-loop model predictive control from Matlab

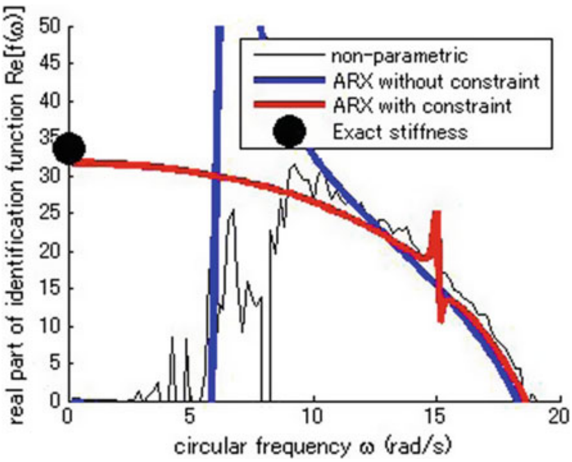


Fig. 13 Multi-loop model predictive control attained from MATLAB simulation by auto regressive partial least squares (ARX)

(i) Without constraint
(ii) With constraint.

The 'K' value is obtained from the mathematical dynamic equation $K = M\omega^2$, for 'M' mass steel frame and 'ω' as frequency (rad/sec).

The stiffness 'K' of steel frame mathematically obtained is compared with experimental test for both simulation and vibration analyses conducted using vibration sensor (Fig. 12).

For this frame analysis, the stiffness of the steel structure (K) due to vibration analysis parameter is computationally predicted with autoregressive exogenous (ARX) (Fig. 13).

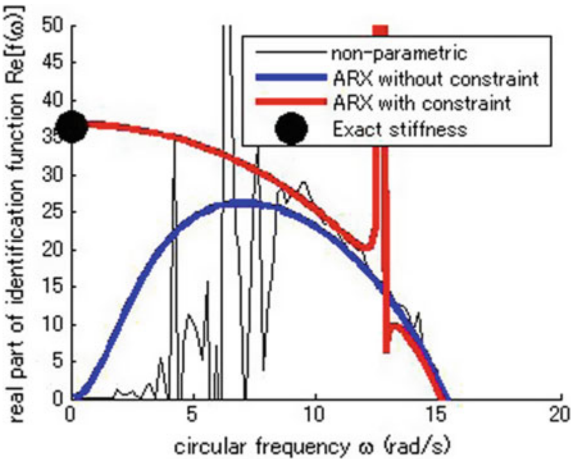


Fig. 14 Auto regressive partial least squares attained from MATLAB using multi-loop model with same frequency

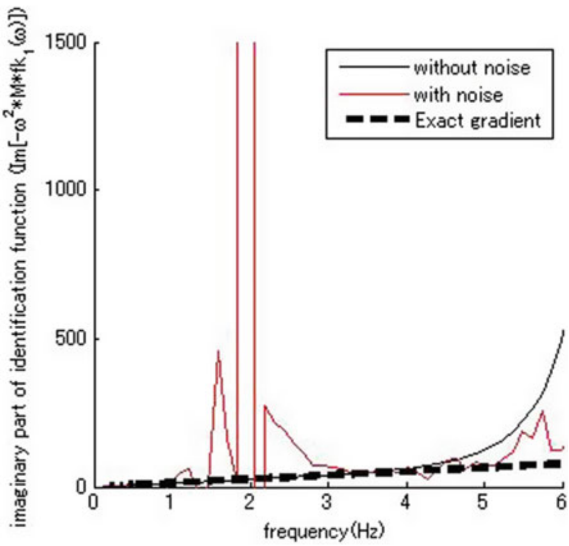


Fig. 15 Frame model vibration analysis with correlation by gradient

For this frame analysis, the stiffness of the steel structure (K) due to vibration analysis parameter is computationally predicted with autoregressive exogenous (ARX) for same frequency (Fig. 14).

For this frame analysis, the stiffness of the steel structure (K) due to vibration analysis parameter is computationally predicted with autoregressive exogenous (ARX) (Fig. 15).

For this frame analysis, exact gradient is taken for plot value as simulation with frequency (Hz) versus stiffness value at without noise condition (Fig. 16).

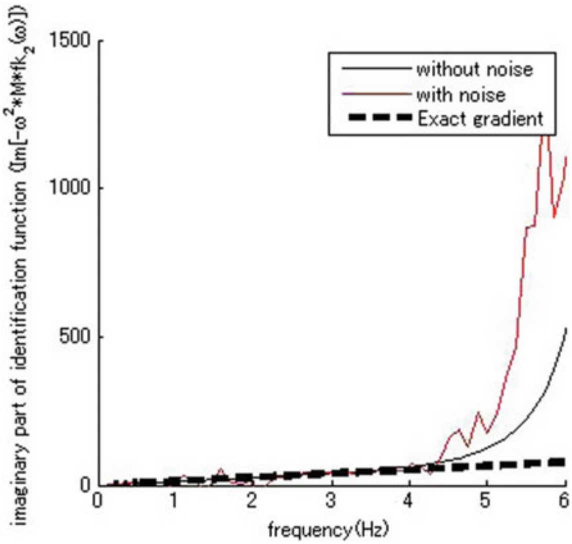


Fig. 16 Vibration analysis for frame model—with noise: real part, without noise: imaginary part correlation by exact gradient

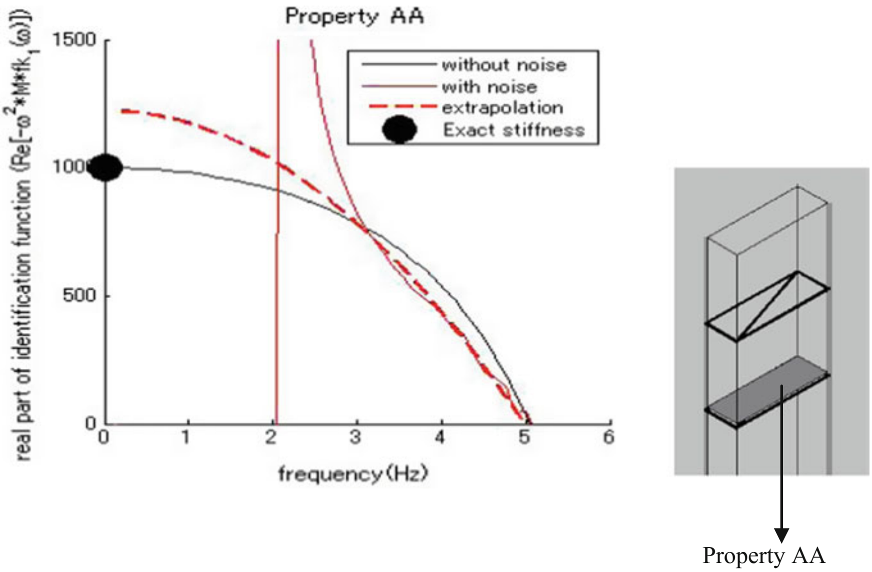


Fig. 17 Property AA—MATLAB attained graph analysis for stiffness (K) Vs Frequency (Hz)

For this frame analysis, exact gradient is taken for plot value as simulation with frequency (Hz) versus stiffness value at without noise condition. Thus, a unique discretization scheme is automatically generated (Fig. 17).

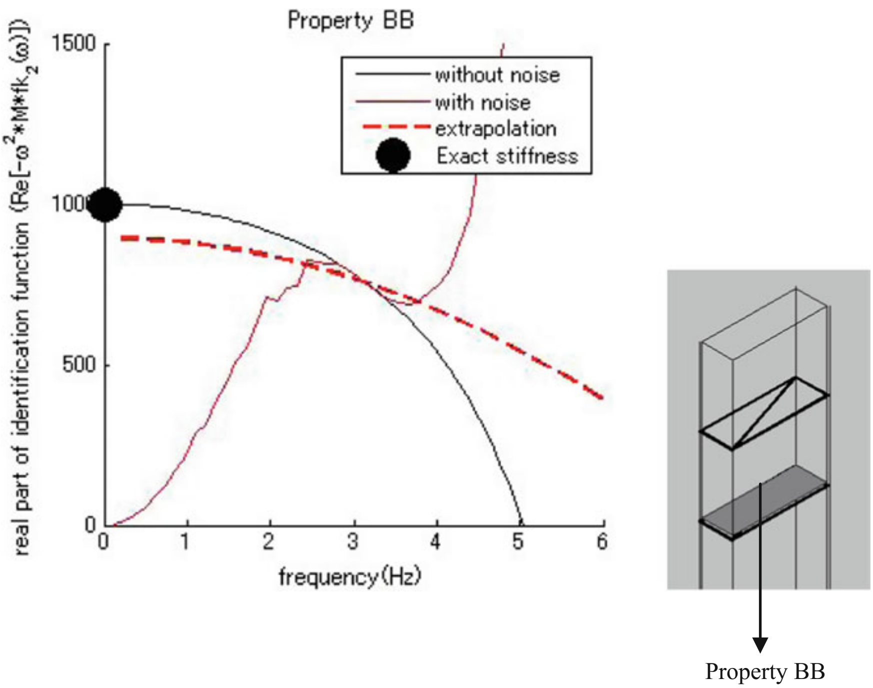


Fig. 18 Property BB—Frame analysis stiffness (K) Vs Frequency (Hz)

From the above-obtained result, the stiffness value for steel frame is same in both the cases of property position of sensors AA and BB with respect to real part of dynamic equation (Fig. 18).

From the above frame positions AA and BB, the analysis is based on the stiffness formulation without noise and with noise condition in different conditions.

From the above condition position of sensor property AA and BB, stiffness (K) of the steel frame obtained and exact stiffness value of steel frame is same in both the experimental and computational cases.

3 Conclusion

In this paper, an experimental test study of vibration analysis is performed on three-storey steel portal frame using vibration motor and computational approach is performed. Frame analysis using Matlab 2013a software and simulation technique is performed and compared with experimental approach done using wireless sensors. In this study, a new simulation algorithm is developed that generates gradient and extrapolation formulation for frame analysis. In this study, **image processing tool** is used for both programming output to Arduino and simulation for Matlab 2013 for analytical and processing approach. Significant future developments of this health monitoring of structures as a multidisciplinary research in structural dynamics, signal processing, motion and smart sensing technologies were discussed. Both experimental works in the lab and computational study were attained and discussed.

References

1. Noel AB, Abderrazak A (2017) Student member structural health monitoring using wireless sensor networks: a comprehensive survey. IEEE Commun Surv Tutorials 19(3)
2. Madhukumar N, Baiju MR (2017) MEMS based wireless sensor network for structural health monitoring. Int JA dv Res Electron Commun Eng (IJARECE) 6(8), ISSN: 2278–909X
3. Lu KC, Wang Y, Lynch JP, Lin PY, Loh C-H, Law KH (2015) Application of wireless sensors for structural health monitoring and control. In: The eighteenth KKCNN symposium on civil engineering december. Taiwan, 19–21
4. Rucka M, Wilde K (2006) Application of continuous wavelet transform in vibration based damage detection method for beams and plates. J Sound Vib 297:536–550
5. Study of SHM from an Integrated approach for structural health monitoring-2010 Prof. Rama shanker department of civil engineering Indian institute of technology Delhi
6. Rizos PF, Aspragathos N, Dimarogonas AD (1990) Identification of crack location and magnitude in a cantilever from the vibration modes. J Sound Vib 138(3):381–388
7. Reda Taha MM, Noureldin A, Lucero JL, Baca TJ ((2005)) Wavelet transform for structural health monitoring: a compendium of uses and features
8. Lynch JP, Wang Y, Sundararajan A, Law KH, Kiremidjian AS (2016) Wireless sensing for structural health monitoring of civil infrastructures
9. Shaladi R, Alasthan F, Yang C (2015) An overview on the applications of structural health monitoring using wireless sensor networks in bridge engineering. In: International conference on advances in science. In: Engineering technology and natural resources (ICASETNR-15) Aug. 27–28, 2015
10. Yan S, Ma H, Li P, Song G, Jianxin W (2017) Development and application of a structural health monitoring system based on wireless smart sensors. Shenyang Jianzhu University, School of Civil Engineering
11. Rogers CA (1990) Intelligent material systems and structures. In: Proceedings of U.S.-Japan workshop on smart/intelligent materials and systems, pp 11–33
12. Rytter A (1993) Vibration based inspection of civil engineering structures. Ph.D. Dissertation, Department of Building Technology and Structural Engineering, Aalborg University, Denmark

An Improved Image Pre-processing Method for Concrete Crack Detection

Harsh Kapadia, Ripal Patel, Yash Shah, J. B. Patel and P. V. Patel

Abstract Structure health monitoring of concrete structures has gained more attention in the recent years due to advancement in the technology. Different methods like acoustic, ultrasonic and image processing based inspection methods have been deployed to carry out an assessment of concrete structure. In this paper, work has been carried out to monitor the health of laboratory scale concrete objects using vision-based inspection. The objective is to provide a modified image pre-processing algorithms for accurate concrete crack detection. Different image processing based algorithms reviewed from existing literature were implemented and tested to detect cracks on the surface of a 15 × 15 × 15 cm concrete cube. Due to random unevenness on the surface of concrete blocks, designing of an accurate and robust algorithm becomes difficult and challenging. Developed algorithm was applied to different images of concrete cubes. Receiver operating characteristics analysis and computation time analysis along with result images were discussed in the paper. In order to validate the applicability of developed algorithm, test results of crack detection on practical crack images are presented. Python was used to develop algorithm along with OpenCV library for image processing functions.

H. Kapadia (✉) · R. Patel · Y. Shah · J. B. Patel · P. V. Patel
Instrumentation and Control Engineering Department, Institute of Technology, Nirma University, Ahmedabad, Gujarat, India
e-mail: harsh.kapadia@nirmauni.ac.in

R. Patel
e-mail: 13bic037@nirmauni.ac.in

Y. Shah
e-mail: 13bic052@nirmauni.ac.in

J. B. Patel
e-mail: jbpatel@nirmauni.ac.in

P. V. Patel
e-mail: paresh.patel@nirmauni.ac.in

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_149

1 Introduction

Image processing involves processing of the digital image data to extract important information and/or features. The extracted information can either be used for autonomous machine perception or human perception or storage or transmission. To extract the information, there are numerous operations, algorithms and methods are available. Researchers around the globe are putting continuous efforts to bring improvements in them. There are numerous applications which cover a wide spectrum of engineering, science, agriculture and technology. In the majority of the cases, an image processing based product is deployed in applications to imitate the human wherein the task performed by human eye and brain is carried out by a camera and a computing device. With the advancement of technology, the prices of a good quality camera and computing devices have decreased to a greater extent. That has enabled people to explore new applications of image processing and improve the performance of the existing systems.

Majority of the inspection applications are carried out by image processing. It involves inspection of variations in a manufactured good against its specifications or may involve laboratory scale inspection of an object under test. Characteristics of the concrete structure are affected by an inferior standard of maintenance, fluctuation of temperature, an excessive amount of load. Cracks are the only external and primary indication of deformities which need to be inspected at earlier stages for strength and reliability of concrete structure [1]. Various kind of cracks is to be identified in distinct types of surfaces like bridges, building, tunnel, road, and pavement [2]. Examination of cracks in practical bases is requisite in a nuclear plant to find out any leakages to prevent a hazardous explosion. Detection and analysis of crack parts on a concrete structure are generally carried out manually in civil engineering which is time-consuming and includes human interference which may lead to high inaccuracy. Normally these inspections are carried out by experts to obtain accurate observation.

Image processing based inspection is useful due to its real-time application, precise accuracy, and cost-effectiveness and fully automated. Ultra-sonic and acoustics wave-based methods are more costly and difficult to implement in practical solution for structure monitoring [3, 4]. Imaging-based inspection approach gives a cost-effective, real-time solutions to monitor structure health [5, 6]. Various type of noises may occur in automatic crack detection such as irregular conditions in illumination, shading, and divots etc. in the images of the concrete structure. Abstruse to discriminate the difference between crack part, noise and non-crack part of the concrete surface.

Civil engineering laboratories perform a compression test, tensile test etc. on various concrete elements like a cube, column, junction etc. It is also evident that one needs to understand the concept and behaviour of different types of concrete in order to use them in practical applications. It is extremely important to assess their effectiveness by casting and testing them in the laboratory. A comparison of different image pre-processing and processing methods for crack detection have been implemented for concrete cubes of $15 \times 15 \times 15$ cm dimension. The motive behind this work is to automate and digitize the concrete cube testing process using the power of vision and computation.

2 Review on Algorithms

Researchers around the globe have explored different image processing based algorithms to address the problems of automatic crack detection. Crack detection methods using image can be based on visible, infrared, ultrasonic, laser, time of flight diffraction, etc. [4]. Different approaches like edge detection, pre-processing methods, fuzzy logic approach, neural networks, subtraction methods, various filter and thresholding methods etc. [7–10] were found in the existing literature. Applicability of these methods was tested on different surfaces of cracks like concrete structures, bridges, pipes, walls, columns, roads, pavements, tunnels etc. Most surfaces whose damage assessment and reliability are critical were considered in the literature.

Yamaguchi et al. had presented a novel approach to detect cracks in concrete surfaces [11]. They proposed an image-percolation model that identifies crack by referring to the connectivity of brightness and shape of the percolated region. They had assumed cracks were composed of thin interconnected textures. The validity of the proposed work was given by using receiver operating characteristics (ROC) analysis by means of experiments on actual concrete surface images. Fujita et al. had presented their work on crack detection [12]. They had proposed pre-processing method to suppress the effect of noises present on the concrete surfaces. An image subtraction method wherein an original image was subtracted from median filtered image to remove non-uniform illumination, shading etc. A hessian matrix based line filter was presented in the paper to emphasize on cracks. Crack part of the images was separated from background using thresholding the image with all values from 0 to 255. A comparison between 256 binary images and manually traced crack image was presented into evaluate the performance of the presented methods. Authors had done receiver operating characteristics analysis for the proposed method and included experimental results. It was concluded that the proposed methods were found effective for crack detection in practical situations. The authors did not mention the type of crack surfaces, brief literature and computation timings of the presented methods.

Atsushi et al. had proposed a crack analysis system for concrete block images. A high-resolution CCD camera was used to acquire crack images [13]. Shading correction, thresholding methods, crack tracking and labelling, feature extraction and analysis were applied to the acquired images and integrated into a tool. The authors had used sub-pixel interpolation method and a standard crack scale in order to measure physical crack area. Two calibration methods were discussed to investigate the efficiency of the proposed method. Further, improvements in the crack area measurement were suggested.

A method of crack detection for the bottom surface of concrete bridges was presented by Xu et al. Image segmentation of crack images was carried out using grey threshold iteration method and canny iteration method to achieve crack extraction [6]. Greying, histogram enhancement and gaussian filtering algorithms were used for pre-processing. Authors had presented a comparison and experimental analysis on the accuracy of two algorithms. In the conclusion, canny iteration method provided better crack extraction. Authors did not include any comparison of computation time.

An overview of different conventional and non-conventional methods of structural health monitoring was discussed by Sharma and Mehta. Variety of other non-destructive evaluation methods based on image processing were discussed along with useful outcomes, constraints, challenges etc. Applications of image processing to monitor structure health has matured and offers non-contact measurement, good amount area, mm accuracy in large structures and micron accuracy in laboratory scale structures [14].

Wang and Huang had summarized crack detection methods based on images and presented the comparison in four categories: integrated algorithm, morphological approach, percolation-based method, and practical technique. Experimental results were presented for each category of the method [10]. It was concluded that integrated algorithm is suitable for shading correction morphological approach gives better results compared to Otsu's threshold, it lacks precision do detect crack pixels. Further, percolation-based method gives best results as it operated on the image locally but the processing time is more compared to other methods. The practical technique is a semi-automatic method which requires human intervention and can give excellent performance for different types of concrete images.

3 Pre-processing Method

In recent times most of the infrastructure is composed of the concrete structure and crack detection is one of the best ways of monitoring health assessment of the structure as crack is one of the primary external, physical and visible parameter on a concrete surface to check the health of a structure. Cracks which appear on a surface are generally random, uneven, can be continuous or broken with random direction. Traditionally the inspection of a structure is carried out by trained professionals.

Vision-based methods can provide an automatic and accurate solution to the problem of automatic crack detection on a concrete surface. Application of vision-based inspection on practical structures is limited due to their large size and limited field of view of a good quality camera. In order to develop automatic crack detection system, laboratory scale concrete specimens were considered in the presented work. Cracks have characteristics like low-intensity values which matches to that of noises/dents on the surface, an area larger than noises/dents but smaller than the background. In the past several methods for crack detection have been proposed on various type of surfaces as discussed above in literature survey. A particular algorithm/method is not applicable for crack detection on all type of concrete surfaces. In this paper, different pre-processing methods were applied and tested on different cases of concrete cubes.

Figure 1 illustrates concrete cube (15 × 15 × 15 cm) image without crack and with cracks.

This paper consists discussion of pre-processing and processing methods as shown in Fig. 2. In order to remove interference objects like noises/dents present on the surface under observation, filter pre-processing was used. Gaussian, averaging, median, min-max etc. filters are widely popular. Bilateral filters can preserve the strong edges while removing the noises. Intensity difference function in bilateral filter only consider pixels with values equivalent to the central pixel value for blurring [15, 16]. Edges remain sharp due to large intensity variation but computation time is slightly higher as compared to other filters like gaussian filter. Image Subtraction is used as a pre-processing method to remove noises and further enable a better differentiation between crack and non-crack objects. Figure 3 shows sample concrete crack images considered for the study.

Fig. 1 Sample image of concrete cube with and without crack

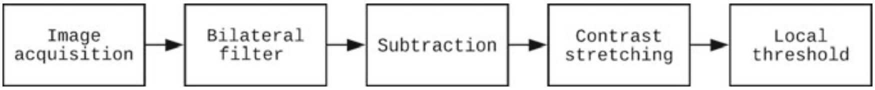


Fig. 2 Flowchart of crack detection

Fig. 3 Sample concrete crack images

3.1 Bilateral Filter Pre-processing

Bilateral filter is highly effective keeping the edges and removing unwanted noise/dents/interference objects etc. Gaussian and other filter considers nearby pixels while filtering, but does not consider the intensity of nearby pixels. Bilateral filter considers both nearby pixels and their intensity while filtering, so as to preserve edges which have large intensity variation [15]. Figure 4 shows the result of bilateral filter operation applied on sample images of Fig. 3. The result indicates that the edges were preserved and noises were removed.

3.2 Subtraction Pre-processing

A synthetic white image was subtracted from the filtered image to obtain a subtracted image. This will remove shadow present on the surface, corrects non-uniform illumination and would give a uniform illuminated image. Normally in vision based inspection system, the enclosed inspection area is provided with constant illumination conditions at the time of image capturing. Due to unsuitable conditions in the inspection, captured images were non-uniform illumination. Thresholding operation [17] was applied to a subtracted image to differentiate crack part from the background. Figure 5 illustrates result image of subtraction pre-processing method applied on sample images of Fig. 3.

Fig. 4 Result images of bilateral filter pre processing

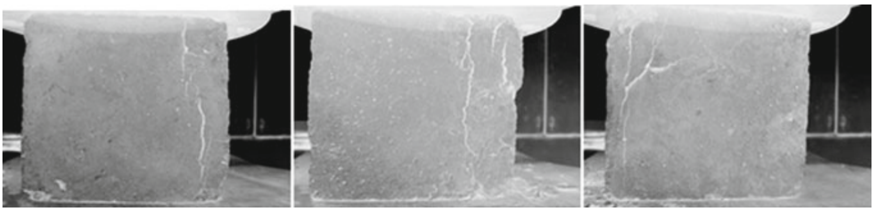


Fig. 5 Result images of subtraction pre processing

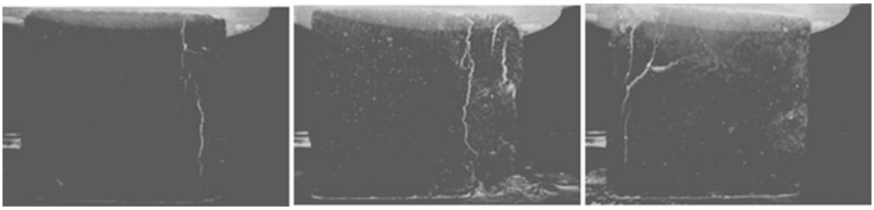


Fig. 6 Result images of contrast stretching

3.3 Contrast Stretching Processing

Contrast stretching method was used to increase the difference between the intensity value of crack part and background. It is evident from the sample images seen in Fig. 1 that intensity values of crack part fall near to intensity values of background and noises due to greyish surface of the concrete cube. This method proves useful as it enhances the contrast of crack pixels and suppresses the non-crack pixels in the images. The range of intensity values of crack pixels was determined and enhance it while changing the pixel values of a non-crack part by multiplying pixel values with zero. This method gives more accurate results when it is applied after subtraction pre- processing method. Shading and non-uniform illumination are corrected by subtraction pre-processing method. Figure 6 shows image obtained after applying contrast stretching on subtracted image applied to sample images of Fig. 3. The result indicates shaded part and noise which are considered as cracks due to the similar intensity values.

3.4 Local Threshold Processing

In order to obtain or carry out crack detection, a threshold operation is required. It will separate crack and background part of the image which can be further utilized to perform crack analysis, measurement of crack length, width etc. and crack characterization. Traditionally, global thresholding like Otsu's thresholding is deployed in cases of bimodal (histogram) image. As already discussed and seen in Fig. 2, concrete surfaces has random noises and dents which hinder accurate crack detection. Local thresholding like niblack or adaptive mean or adaptive gaussian suits such cases [17, 18]. Here, niblack based local thresholding was used and the results of same can be seen in Fig. 7.

Fig. 7 Result images of local threshold

4 Experimental Results and Analysis

Discussed methods were tested for crack detection on more than 50 images similar to the images seen in Figs. 1 and 3. All sample images which were captured by DFK72AUC02 CMOS USB2.0 industrial camera by The Imaging Source GmbH, Germany equipped with 8 mm C-mount lens by Computar, Japan. Camera has sensor size 1/1.25" with micron CMOS sensor (MT9P031) and a rolling shutter [19]. It has a maximum resolution of 5 MP and 6 fps. As it can be seen from above results that most of the crack parts are identified from sample image (Fig. 3) with the combinational use of two pre-processing and two processing methods. Receiver Operating Characteristics (ROC) analysis presented in Table 1 is carried out to evaluate the performance of methods on images of chosen 7 sample images of size 2048 × 1536 (3,145,728 pixels). Crack detection results of the chosen sample images were visually better in terms of crack detection. True positive fraction (TPF) is the ratio of no. of crack pixels in the processed image to the no. of crack pixels in ideal image. It signifies the ability of the method/process to be able correctly identify cracks in the image. Process/method should also correctly to identify background as background which is evaluated false positive fraction (FPF). It is the ratio of no. of background pixels in processed image to the no. of background pixels in the ideal image. TPF should be near to 1 and FPF should be near to 0 which is observed in Table 1 and chart shown in Fig. 8. National instruments LabVIEW vision assistant tool (evaluation version) was utilized to obtain number of crack pixels and number of background pixels in both ideal and processed image. The data mentioned in the Table 1 were obtained using area parameter of particle analysis function available in vision assistant tool. Computational time of methods tested on Fig. 3 is presented in Table 2. It gives a measure of real-time implementation capability of the process/method discussed. As a bilateral filter, contrast stretching and niblack threshold methods are both local processing methods, software code requires more computation time compared to that of global processing methods. In Table 2 all measurement unit of time are in seconds.

Further testing of this methods are carried out on a different set of images which consists various type of surfaces and these images are captured by smartphone. Most of the crack parts are identified easily if there is a wide difference in the contrast values of crack parts and background. Figure 9 row 1 shows sample image taken from the white wall of a building, row 2 is a sample image of the crack surface taken

Table 1 ROC analysis of selected images

No. of crack pixels in ideal image	No. of crack pixels in processed image	True positive fraction (TPF)	No. of background pixels in ideal image	No. of background pixels in processed image	False positive fraction (FPF)
21,806	18,175	83.34	3,123,922	3,127,553	16.65
28,697	25,667	89.44	3,117,031	3,120,061	10.55
19,863	16,101	81.06	3,125,865	3,129,627	18.93
20,303	18,130	89.29	3,125,425	3,127,598	10.70
9766	8200	83.96	3,135,962	3,137,528	16.03
5738	4744	82.67	3,139,990	3,140,984	17.33
7543	5945	78.81	3,138,185	3,139,783	21.19

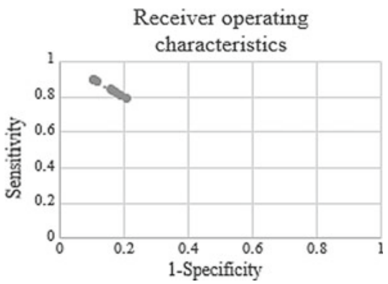


Fig. 8 ROC analysis of selected images

Table 2 Algorithm computation time for different image sizes

Method\image	Set 1 (1296 × 972)	Set 2 (2048 × 1536)	Set 3 (2592 × 1944)	Set 4(4032 × 3024)
Filter + Subtraction + Contrast stretching + Threshold	0.16	0.40	0.62	1.45

from the corridor which consists slightly uneven surface and row 3 is sample image taken from a road which has high unevenness and low contrast difference between crack and non-crack part. Figure 10 shows the threshold result of each image in the column. It is clearly seen that combination of subtraction pre-processing and contrast stretching method is effective to get more accurate crack detection results on the majority of the surfaces. The method can be very effective to find a crack if there is a significant difference of intensity value between foreground and background but for images with low contrast different like road surface shown in 3rd column of Fig. 10 accuracy is not very high.

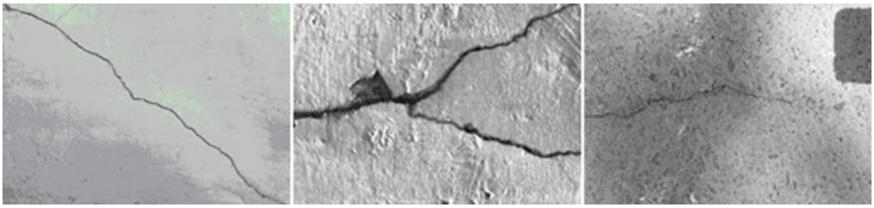


Fig. 9 Sample concrete crack images

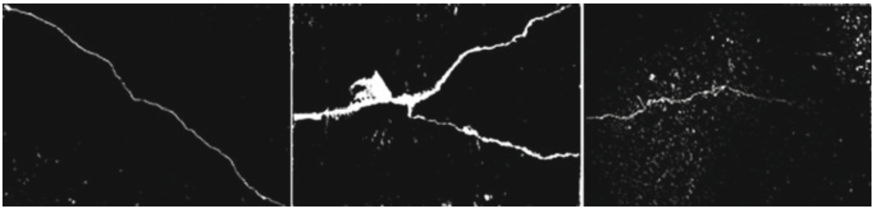


Fig. 10 Threshold images of sample images shown in Fig. 9

5 Conclusion

In this paper, two pre-processing and two processing methods are implemented for improved crack detection on different images on various crack surfaces. The methods were found effective when it is implemented on images with high contrast difference between crack part and background. Random unevenness on the surface of concrete cube presents challenges in the development of robust image processing algorithm for accurate crack detection. The proposed method detects cracks on the surface more accurately compared to the other methods suggested in the literature. Also, it is evident from the results that, a crack detection algorithm designed to identify cracks on the surface of a bridge will not provide similar accurate results when evaluated for different surfaces of concrete containing crack. The developed algorithm gives accurate results on the majority of the crack containing concrete surfaces as indicated in the Fig. 10 and presented ROC analysis. Results show that computation time increases with an increase in the size of an image. It is evident from the timings that, the algorithms can be useful for online/real-time crack detection implementation. A graphical user interface tool using python and OpenCV [20] can be developed to provide an open source crack analysis tool. It will allow engineers/students/researchers to carry on crack detection and analysis on different concrete surfaces.

References

1. Cheng C-C, Cheng T-M, Chiang C-H (2008) Defect detection of concrete structures using both infrared thermography and elastic waves. Autom Concr, pp 87–92

2. Adhikari RS, Moselhi O, Bagchi A (2014) Image-based retrieval of concrete crack properties for bridge inspection. Autom Constr 39:180–194
3. Koch Christian, Georgieva Kristina, Kasireddy Varun, Akinci Burcu, Fieguth Paul (2015) A review on computer vision based defect detection and condition assessment of concrete and asphalt civil infrastructure. Adv Eng Inform 29(2):196–210
4. Mohan A, Poobal S (2017) Crack detection using image processing: a critical review and analysis. Alexandria Eng J
5. Lee BY, Kim YY, Yi S-T, Kim J-K (2013) Automated image processing technique for detecting and analyzing concrete surface cracks. Struct Infrastruct Eng 9(6):567–577
6. Xu H, Tian Y, Lin S, Wang S (2013) Research of image segmentation algorithm applied to concrete bridge cracks. In: International conference on information science and technology (ICIST). IEEE, pp 1637–1640
7. Talab AMA, Huang Z, Xi F, HaiMing L (2016) Detection crack in image using Otsu method and multiple filtering in image processing techniques. Optik-Int J Light Electron Opt 127(3):1030–1033
8. Miyamoto A, Konno M-A, Bruhwiler E (2007) Automatic crack recognition system for concrete structures using image processing approach. Asian J Inf Technol Medwell J 6(5)
9. Choudhary GK, Dey S (2012) Crack detection in concrete surfaces using image processing, fuzzy logic and neural network. In: IEEE fifth international conference on advanced computational intelligence (ICACI). Nanjing, Jiangsu, China
10. Wang P, Huang H (2010) Comparison analysis on present image-based crack detection methods in concrete structure. In: 3rd international congress on image and signal processing (CISP). IEEE, pp 2530–2533
11. Yamaguchi T, Nakamura S, Saegusa R, Hashimoto S (2008) Image based crack detection for real concrete surfaces. IEEJ Trans Electr Electron Eng 3(1):128–135
12. Fujita Y, Mitani Y, Hamamoto Y (2006) A method for crack detection on concrete structure. In: 18th international conference on pattern recognition ICPR. IEEE, pp 901–904
13. Atsushi I, Aoki Y, Hashimoto S (2002) Accurate extraction and measurement of fine cracks from concrete block surface image. In: Proceedings of the 2002 28th annual conference of the ieee industrial electronics, pp 2202–2207
14. Sharma A, and Mehta N (2016) Structural health monitoring using image processing techniques-a review. Int J Mod Comput Sci 4(4):93–97
15. Tomasi C, Manduchi R. (1998) Bilateral filtering for gray and color images. In: Sixth international conference on computer vision, IEEE. pp 839–846
16. 2015. Smoothing Images. December 18. Accessed October 20, 2017. https://docs.opencv.org/3.1.0/d4/d13/tutorial_py_filtering.html
17. Niblack and Sauvola Thresholding. Accessed March 22, 2018. http://scikit-image.org/docs/dev/auto_examples/segmentation/plot_niblack_sauvola.html
18. Saxena LP (2017) "Niblack's binarization method and its modifications to real-time applications: a review. Artif Intell Rev 1–33
19. DFK 72AUC02 USB2.0 CMOS color industrial camera. Accessed November 2, 2017. https://www.theimagingsource.com/products/industrial-cameras/usb-2.0-color/dfk72auc02/
20. 2014. OpenCV-Python Tutorials. November 14. Accessed December 12, 2017. https://docs.opencv.org/3.0-beta/doc/py_tutorials/py_tutorials.html

Grape Crop Disease Classification Using Transfer Learning Approach

K. R. Aravind, P. Raja, R. Aniirudh, K. V. Mukesh, R. Ashiwin and G. Vikas

Abstract Grape is one of the important fruit crops which is affected by diseases. The advents of digital camera and machine learning based approaches have facilitated recognition of plant diseases. Convolution Neural Network (CNN) is one of the types of architecture used in deep learning based approach. AlexNet is a category of CNN which is used in this study for classification of three diseases along with healthy leaf images obtained from PlantVillage dataset. Transfer learning-based approach is used where the pretrained AlexNet is fed with 4063 images of above categories. The model achieved 97.62% of classification accuracy. Feature values from the different layers of the same network are extracted and applied to Multiclass Support Vector Machine (MSVM) for performance analysis. Features from Rectified Linear unit (ReLu 3) layer of AlexNet applied to MSVM achieved the best classification accuracy of 99.23%.

1 Introduction

Agricultural crops are threatened by the incidence of pests and diseases which in turn affect its production. The total loss of global production due to plant diseases is roughly estimated to be 10% [1]. Plant diseases are mainly caused by three pathogenic agents, namely, fungi, bacteria, and viruses [2]. Three diseases of grape crop, namely, black rot, black measles, and Isariopsis leaf spot, are considered in this study. Grape is one of the economically important crops with worldwide production of 7.8 million tons (2016) of which 39% are from Europe, 34% are from Asia, and 18% are from America [3]. The crop's production is affected widely by abovementioned diseases. Black rot is one of the deadly diseases caused by fungus *Guignardia bidwellii* which sometimes can cause 100% crop loss [4]. Black measles also called Esca caused by fungi are widespread in Europe which cause significant damage to the crop [5].

K. R. Aravind · P. Raja (✉) · R. Aniirudh · K. V. Mukesh · R. Ashiwin · G. Vikas
School of Mechanical Engineering, SASTRA Deemed University, Thanjavur, India
e-mail: raja@mech.sastra.edu

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_150

Isariopsis leaf spot caused by fungi *Isariopsis clavispora* leads to serious damage to the crop under favorable climatic conditions [6].

Traditionally, plant diseases are diagnosed by experts through visual inspection which is prone to errors [2, 7]. Advances in science and technology are providing better tools and techniques for early detection, prevention, and control of pests and diseases [1]. Recent revolution in smartphones and digital cameras are paving a way to utilize these cheap image acquisition devices directly or by integration to an automation device for plant disease recognition applications [8]. Many studies have been carried out for foliage disease recognition using image processing techniques and machine learning algorithm in a variety of crops [8–21].

The recognition process involves classification of symptomatic patterns on the surface of leaves for identification of specific foliage diseases. Traditionally, disease recognition using image processing follows several steps such as preprocessing, color transformation, segmentation, feature extraction, and classification [12]. Dey et al. [13] presented a method where the initial image is cropped to reduce Central Processing Unit (CPU) processing time and storage. The image was transformed from Red Green Blue (RGB) color space to Hue Saturation Value (HSV). A thresholding operation based on Otsu method was carried out in hue channel of the HSV image for segmentation, and rotten area of the affected leaf was identified. In a study, Arivazhagan et al. [10] demonstrated a method where initially, image from Red Green Blue (RGB) space is converted to Hue Saturation Intensity (HSI) space and segmentation was carried out using thresholding operation. Textural features such as energy, homogeneity, contrast, cluster shade, etc. were extracted and provided as input to Support Vector Machine (SVM) which is used widely for classification problem. Qin et al. [11] presented a method where the pixels in the image were clustered using *K* median and fuzzy C-means clustering algorithms into ten classes. Based on the mean of hue component and application of classification algorithm (Naive Bayes, logistic regression, and linear discriminant analysis) using a^* and b^* values, the pixels are classified as a part of a lesion or healthy region. Among the total 129 extracted features, best features were selected using different methods (1-rule, correlation-based feature selection, and ReliefF method), and results were compared. In all the above-discussed studies, features of interest are extracted and tested for its accuracy. The accuracy of classification depends on the selected features and it is a time-consuming process.

Recent advances in the development of complex architectures using artificial neural network laid foundation for deep learning based classification. Different approaches are followed for building the network which resulted in many deep learning based algorithm, namely, autoencoder, sparse coding, Convolutional Neural Networks (CNN), and restricted Boltzmann machines. Among these algorithms, CNN is widely used for vision-based application [22]. AlexNet, GoogLeNet, Visual Geometry Group (VGG), etc. are some of the examples of CNN-based architectures. Few studies have already explored the use of CNN in plant disease classification [8, 14, 18]. In this study, three diseases of grape crops along with healthy leaves are classified using the pretrained AlexNet framework which is known as transfer learning. In addition, feature values are extracted from the different layers of a pretrained AlexNet framework and provided as input to the Multiclass Support Vector Machine (MSVM). The classification performance of the features from each layer is compared and analyzed.

2 Method

2.1 Dataset

The segmented image datasets for the three diseases, namely, black measles, black rot, and Isariopsis leaf spot (as shown in Fig. 1), along with the healthy leaves are obtained from PlantVillage dataset which consists of 54,306 images [23]. The obtained subset consisting of three diseases and healthy leaf images constitutes 4,063 segmented images.

2.2 Implementation

AlexNet is a category of CNN which consist of five convolution layers, seven rectified linear unit (ReLu), three max pooling layers, three fully connected layers, two normalization layers, two drop out layers, and one softmax layer as shown in Fig. 2. The network was pretrained with ImageNet dataset and participated in the ImageNet Large-Scale Visual Recognition challenge (ILSVR—2010) with 1.2 million images of over 1000 categories. The network ability for classification was tested and proved to be efficient than the previous benchmark. A variant of the AlexNet was implemented in ILSVR—2012 resulting in better error rates with its implementation [24].

Training of these networks with the new class of images for disease recognition requires a large dataset and time-consuming process. Hence, transfer learning tech-

(a). Black measles (b). Black rot (c). Isariopsis leaf spot

Fig. 1 Segmented image of the grape crop leaves from PlantVillage dataset

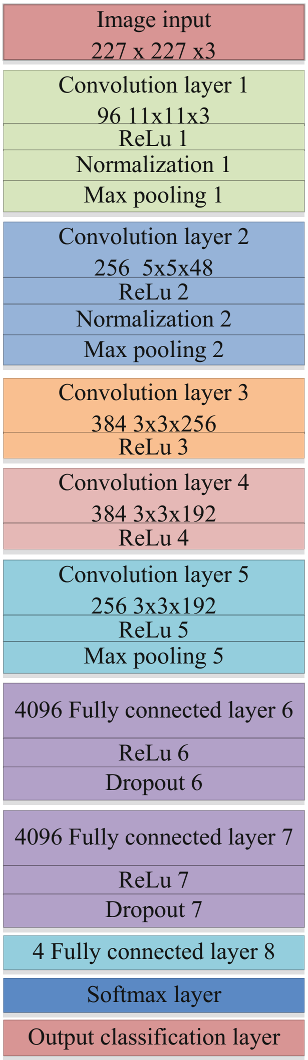


Fig. 2 AlexNet architecture implemented for this study

nique is used where the AlexNet is pretrained with ImageNet. The features or layer weights learned from the pretraining process is preserved for its application in disease recognition process. The size of the segmented image from the dataset is 256×256 which was resized to 227×227 as the input layer of AlexNet in Matlab can accept image in the above-specified dimension. The convolution layers 1 and 2 is followed by ReLu, normalization, and max pooling. ReLu is a nonlinearity-based activation function implemented in the output of all convolution layers. Cross-channel normalization is performed to amplify the output of the excited neurons which have higher activation and suppress the output of the local neighborhood neurons. Max pooling is done to reduce the dimension of the output from the previous convolution layers by finding the maximum value within the kernel applied to the output of the convolutional layer. The output from the convolution layer 3 and 4 is fed to the ReLu, while output from layer 5 is provided to both ReLu and max pooling. The output from the previous layer is fed to the two 4096 fully connected layers where all the neurons are connected to each other. In dropout layers 50% of the neurons are disconnected randomly from the adjacent neurons momentarily and training is carried out using the resulting network. This method has found to lower the generalization errors. Originally, AlexNet was trained to classify 1000 categories of object; hence, it had 1000 fully connected layer. In this study, it has been replaced with four fully connected layers as only four categories are used. The softmax layers provide the probability value of the given sample image to all the class labels. The label with the highest probability is the correct class label.

The AlexNet consists of 23 layers excluding the input and the output layers. Feature values from each layer are extracted and fed as input to the MSVM. The MSVM model was implemented with "one versus all" approach where the multiclass problem was split into several binary classification problems. The result for the features from each layer is compared and discussed in Sect. 3.

2.3 Visualization of the Activation Map

In an artificial neural network with few layers and neurons, the features are extracted from the image and fed as input to the network. Contrastingly, in AlexNet, a category of CNN automatically learns to identify features. The AlexNet have five convolution layers, wherein each layer, the convolution operation is performed using a kernel matrix of specific dimension and applied over the input image. The applied matrix moves over the entire image, and an activation map or feature map is obtained as a result of the convolution operation in each convolution layer.

The number of kernel matrix and its dimension varies for each layer as shown in Fig. 2. For example, the convolution layer 1 consists of 96 kernel matrix with a dimension of $11 \times 11 \times 3$. Hence, 96 different channels of activation maps are obtained for the given input image by passing it through the first convolution layer with a total of 290,400 values. The activation map shows the specific region for which the neuron becomes active while applying the kernel over the image. The activation map can be visualized using the "montage" function in Matlab after normalization of the values. An activation map generated from the sample image of black measles is shown in Fig. 3. The activation maps are reshaped to an 8×12 grid for presenting all the 96 channels in a single image. Each channel shows a region with white, black, and gray color.

The region with white denotes positive activation, while black depicts negative activation. The region with gray is positive but not as strong as the region with white color. In Fig. 3, region with disease symptoms has high positive activation in 10th (as shown in Fig. 4) and 32nd channels, and healthy region has high negative activation,

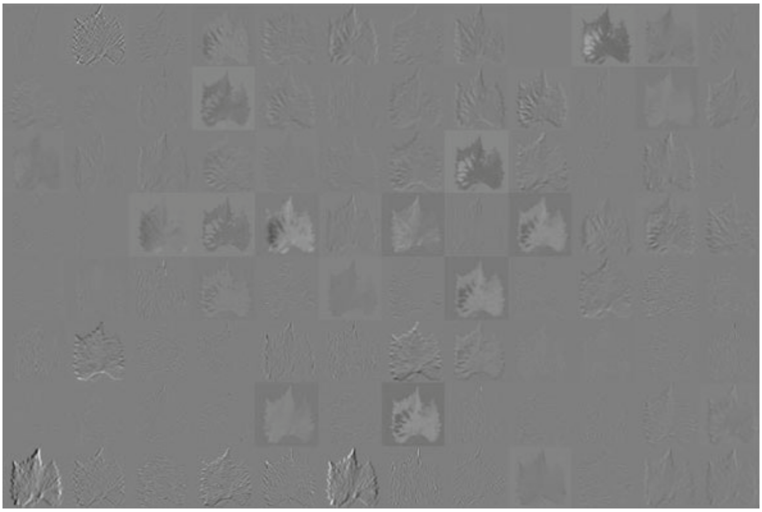


Fig. 3 Features map obtained from convolution layer for black measles

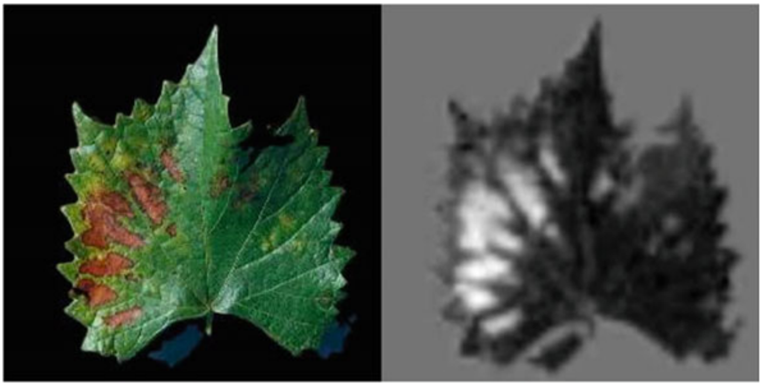


Fig. 4 Comparison of the activation channel 10 with RGB image

whereas in 41, 45, 56, and 78th channels, it has opposite effect in activation. Figure 4 shows the comparison of activation map of 10th channel and RGB image.

The activation map obtained from convolution layer acts as the input to the second convolution layer after passing through ReLu, normalization, and max pooling layers. The convolution operation is performed using 256 kernels with the dimension of $5 \times 5 \times 48$ in the second layer. Certain mid-level features are obtained in the second layer. Subsequent pass through the other layer will result in activation maps which combines the low-level features obtained from the other layers into complex features. The features from these various layers are obtained, and its accuracy in classification using MSVM for the given set of images are analyzed which is discussed in next section (Sect. 3). Also, the accuracy resulting from AlexNet using transfer learning is discussed.

3 Results and Discussions

The system used for the study have GTX 1050 4 GB Graphics Processing Unit (GPU) with 8 GB Random Access Memory (RAM). The segmented image of dimension 227 × 227 from the dataset is fed to the AlexNet. 80% of the images from each class label were used for training and 20% were used for testing. Table 1 shows the number of images present in each class label.

The weight and bias learn rate of all the previous layers are not changed except the last fully connected layers. The weight learn rate is set to 10 and bias learn rate to 20. This means that the weight and bias learn rate are 10 and 20 times faster than the global learning rate, respectively. The maximum number of epoch is set as the default value 10, and a mean accuracy of 97.62% is obtained as shown in Table 2. The number of feature values obtained from each layer is shown in Table 3.

The number of feature values of ReLu layer is same as the preceding convolution layer. The dimension of features reduces as it passes through the subsequent convolution layers in the AlexNet. The accuracy of the classification using MSVM by obtaining the feature values from the five convolution layers is shown in Fig. 5. The best mean accuracy of 99.16% was obtained using the features obtained from the convolution layer 4. Feature values from the convolution layer 1 resulted in a poor accuracy of 82.61% as only simple features such as edges, color information, etc. are captured which is not sufficient for classification. The performance of the values obtained from the other layers (as shown in Table 4) was also analyzed.

Table 1 Grape leaves dataset

Class label	No. of images
Black rot	1180
Black measles	1384
Healthy	423
Isariopsis leaf spot	1076
Total	4063

Table 2 Accuracy of AlexNet in classification

Accuracy				
Black rot (%)	Black measles (%)	Isariopsis leaf spot (%)	Healthy (%)	Mean (%)
94.07	95.31	99.53	100	97.62

Table 3 Number of features from each layer

Layers	No. of feature values
Convolution layer 1	290,400
Convolution layer 2	186,624
Convolution layer 3	64,896
Convolution layer 4	64896
Convolution layer 5	43,264
Max pooling layer 1	69,984
Max pooling layer 2	43,264
Max pooling layer 5	9216
Drop 6 and 7	4096

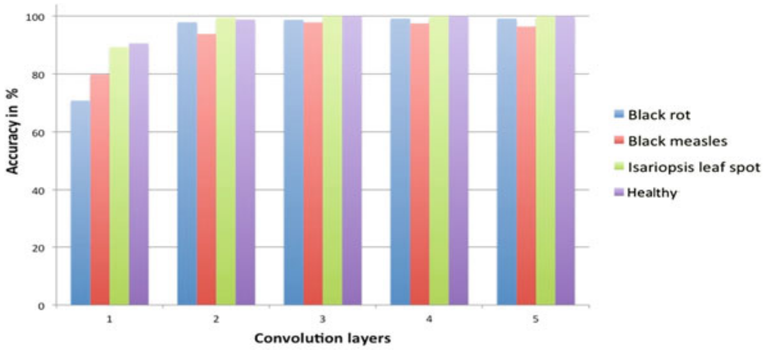


Fig. 5 Accuracy of classification for features from convolution layer 1–5

The feature values obtained from ReLu 3 resulted in the best mean accuracy of 99.23%. The mean classification accuracy has improved significantly when feature values from the different layers of the AlexNet are extracted and applied to shallow classification algorithm such as MSVM. AlexNet performance was better when compared with traditional approaches such as the study by Wang et al. [25] where two kinds of grape disease, namely, grape powdery mildew and Downy mildew with a limited dataset, were classified. The prediction accuracy was 100% using Radial Basis Function (RBF) neural network while with Backpropagation networks (BP), generalized regression networks, and Probabilistic Neural Network (PNN) yielded 94.29% of accuracy. Four shape features, 21 color features, and 25 textural features were extracted for classification, whereas in our case the features are extracted automatically. The number of dataset used in the study is relatively low of about 50 images for each category, but in our case 4063 images of standard PlantVillage dataset have been used. In another study, scab and rust of grape crop have been recognized using a total of 1478 images with Gabor filter-based features and color features. MSVM have been used for classification with an average accuracy of 86.03% which is lower

Table 4 Accuracy of classification using features from different layers

Layers	Black rot (%)	Black measles (%)	Isariopsis leaf spot (%)	Healthy (%)	Mean accuracy (%)
ReLu 1	87.29	90.97	97.21	97.65	93.28
Normalization 1	89.41	93.86	98.14	100	95.35
Max pooling 1	97.46	93.50	99.53	100	97.62
ReLu 2	97.46	94.95	100	100	98.10
Normalization 2	98.31	95.67	100	100	98.49
Max pooling 2	99.15	97.47	100	100	99.16
ReLu 3	98.73	98.19	100	100	99.23
ReLu 4	98.73	97.47	100	100	99.05
ReLu 5	97.03	97.83	100	100	98.72
Max pooling 5	97.88	98.19	100	100	99.02
4096 fully connected layer 6	96.61	97.11	100	100	98.43
ReLu 6	97.88	97.11	100	100	98.75
Dropout 6	97.88	97.11	100	100	98.75
4096 fully connected layer 7	96.61	97.11	100	100	98.43
ReLu 7	98.73	97.11	100	100	98.96
Dropout 7	98.96	97.11	100	100	98.96

than the accuracy discussed in this study [26]. Therefore, by extracting features from different layers of pretrained AlexNet and applying it to MSVM increases the accuracy significantly which is evident from the results. The results also suggest that deep learning methods are better than the methods where feature is extracted manually and applying it to the shallow learning algorithms.

4 Conclusion

Deep learning approaches require large dataset for training and results in better classification. When the number of the dataset is relatively low, a pretrained network can be used for new classification problem which is known as transfer learning. Further, the accuracy can be improved by extracting feature values from the trained AlexNet and applying it to the shallow learning based classification algorithm. The

AlexNet yielded an accuracy of 97.62%, while the classification with MSVM using feature values from different layers resulted in improved performance. Feature values from ReLu 3 resulted in best performance of 99.23% accuracy. Further evaluation can be done by using different datasets and machine learning algorithms, expanding it to other disease recognition applications.

References

1. Strange RN, Scott PR (2005) Plant disease: a threat to global food security. Phytopathology 43:83–116
2. Gavhale KR, Gawande U (2014) An overview of the research on plant leaves disease detection using image processing techniques. IOSR J Comput Eng 16(1):10–16
3. World Vitinviniculture Situation, OIV Statistical Report on World Vitinviniculture, International Organisation of Vine and Wine. www.oiv.int/public/medias/5479/oiv-en-bilan-2017.pdf
4. Sosnowski MR, Emmett RW, Wilcox WF, Wicks TJ (2012) Eradication of black rot (*Guignardia bidwellii*) from Grapevines by drastic pruning. Plant Pathol 61:1093–1102
5. Mugani L, Graniti A, Surico G (1999) Esca (Black Measles) and brown wood-streaking: two old and elusive diseases of Grapevines. Plant Dis 83:5
6. Silva CMD, Estrada KRFS, Rios CMFD, Batista BN, Pascholati SF (2014) Effect of culture filtrate of *Curvularia inaequalis* on disease control and productivity of Grape cv. Isabel. Afr J Agric Res 9(40):3001–3010
7. Garcia J, Barbed A (2016) A review on the main challenges in automatic plant disease identification based on the visible range images. Biosys Eng 144:52–60
8. Mohanty SP, Hughes DP, Salathe M (2016) Using deep learning for image-based plant disease detection. Front Plant Sci 7, Article ID 1419
9. Hiary HA, Ahmar SB, Reyalat M, Braik M, Rahamneh ZA (2011) Fast and accurate detection and classification of plant diseases. Int J Comput Appl 17(1):31–38
10. Arivazhagan S, Shebiah RN, Ananthi S, Varthini SV (2013) Detection of unhealthy region of plant leaves and classification of plant leaf diseases using textural features. Agric Eng Int CIGR J 15(1):211–217
11. Qin F, Liu D, Sun B, Ruan L, Ma Z, Wang H (2016) Identification of Alfalfa leaf diseases using image recognition technology. PLoS ONE 11(12):e0168274
12. Singh V, Misra AK (2017) Detection of plant leaf diseases using image segmentation and soft computing techniques. Inf Process Agric 4:41–49
13. Dey AK, Sharma M, Meshram MR (2016) Image processing based leaf rot disease, detection of Betel Vine (*Piper BetleL.*). Procedia Comput Sci 85:748–754
14. Sladojevic S, Arsenovic M, Anderla A, Culibrk D, Stefanovi D (2016) Deep neural network based recognition of plant diseases by leaf image classification. Comput Intell Neurosci 2016, Article ID 3289801
15. Barbedo JGA, Koenigkan LV, Santos TT (2016) Identifying multiple plant diseases using digital image processing. Biosys Eng 147:104–116
16. Mahlein AK (2016) Plant disease detection by imaging sensors—parallels and specific demands for precision agriculture and plant phenotyping. Plant Dis 241–251
17. Zhang S, Wu X, You Z, Zhang L (2017) Leaf image based cucumber disease recognition using sparse representation classification. Comput Electron Agric 134:135–141
18. Ferentinos KP (2018) Deep learning models for plant disease detection and diagnosis. Comput Electron Agric 145:311–318
19. Ali H, Lali MI, Nawaz MZ, Sharif M, Saleem BA (2017) Symptom based automated detection of citrus diseases using color histogram and textural descriptors. Comput Electron Agric 138:92–104

20. Zhang S, Wang H, Huang W, Zhuhong Y (2018) Plant disease leaf segmentation and recognition by fusion of superpixel, K-means and PHOG. Optik 157:866–872
21. Xu P, Wu G, Guo Y, Chen X, Yang H, Zhang R (2017) Automatic wheat leaf rust detection and grading diagnosis via embedded image processing system. Procedia Comput Sci 107:836–841
22. Guo Y, Liu Y, Oerlemans A, Lao S, Wu S, Lew SL (2016) Deep learning for visual understanding: a review. Neurocomputing 187:27–48
23. Hughes DP, Salathe M. An open access repository of images on plant health to enable the development of mobile disease diagnostics. arXiv:1511.08060
24. Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural network. In: Advances in neural information processing systems. Lake Tahoe, pp 1097–1105
25. Wang H, Li G, Ma Z, Li X (2018) Application of neural networks to image recognition of plant diseases. In: International conference on systems and informatics. Yantai, pp 2159–2164
26. Meunkaewjinda A, Kumsawat P, Attakitmongcol K, Srikaew A (2008) Grape leaf disease detection from color imagery using hybrid intelligent system. In: International conference on electrical engineering/electronics, computer, telecommunications and information technology. Krabi, pp 513–516

Exploring Image Classification of Thyroid Ultrasound Images Using Deep Learning

K. V. Sai Sundar, Kumar T. Rajamani and S. Siva Sankara Sai

Abstract Deep learning for medical imaging has been at the forefront of its numerous applications, thanks to its versatility and robustness in deployment. In this paper, we explore various classification methodologies that are employed for datasets of relatively small in size to actually train a deep learning algorithm from scratch. Thyroid ultrasound images are classified using a small CNN from scratch, transfer learning and fine-tuning of Inception-v3, VGG-16. We present a comparison of the aforementioned methods through accuracy, sensitivity, and specificity.

1 Introduction

1.1 Background

Image classification task of a machine learning/deep learning algorithm is widely found to be useful in the medical imaging domain, wherein the algorithm is trained to classify medical images into benign or malignant in case of cancer detection and various other ailments based on the symptoms in the images. Arriving at automated diagnosis seems to be the dream of every researcher associated with computer vision coupled with biomedical imaging. Contrary to famous datasets like the ImageNet dataset, Cifar-10, and Cifar-100 which have thousands of images in each category, the availability of medical data is limited. When running a deep learning algorithm with millions of parameters, less data hurdles the training with overfitting. The model eventually tends to fail at generalization of learning giving low accuracy on the test dataset. Regularization can reduce the high variance to some extent but training a deep learning framework from scratch remains out of bounds. Data augmentation

K. V. Sai Sundar (✉) · S. Siva Sankara Sai
Sri Sathya Sai Institute of Higher Learning, Puttaparthi, India
e-mail: saisundarkandarpa@gmail.com

K. T. Rajamani
Robert Bosch, Bangalore, India

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_151

can be employed to further boost the dataset size. Therefore, in order to arrive at accuracies which can be of deployment standard, we resort to training a smaller CNN from scratch, transfer learning—using bottleneck features from deep CNNs to train a new FC layer or a different classifier and finally fine-tuning deep architectures to classify the custom dataset.

1.2 Motivation

The thyroid ultrasound domain was chosen following consultation with local doctors who brought to our attention the high prevalence of thyroid ailments in the region. Thyroid ultrasound is predominantly used to detect thyroid nodules, classify them as benign or malignant and also to identify goiter, thyroiditis. The problem consists of binary classification initially identifying images of patients who probably require a biopsy in order to confirm malignancy of nodule and eventually multi-class classification identifying various other ailments apart from cancer. The aim of this work is to develop an automated thyroid diagnosis system that could aid the radiologist and fasten the diagnosis. In this paper, we only discuss our implementations of binary classification.

2 Dataset Description

We have used two datasets in this work. The first dataset [1] is a publically available one consisting of 298 images and their corresponding biopsy verified reports in.xml format. The TIRADS scores are given for each of the images ranging from 2 to 5 on the scale of increasing probability for malignancy. Since our task in this work dealt only with probably benign or malignant test scenario, we considered scores 2 and 3 as benign and all the scores above these as malignant. The second dataset used in this work was the local database of images from GE LOGIQ P9 which were labeled by an experienced doctor and the reports were written in word format. The various cases of cancerous nodule, thyroiditis, simple goiter, multinodular goiter, toxic goiter, and normal were present in this database. Again, we considered only the relevant images as mentioned for the previous dataset. This dataset consisted of thyroid images of 127 patients.

3 Frameworks and Hardware Used

The deep learning frameworks TensorFlow [2] and Keras [3], based in Python have been used in this work. TensorFlow, released by Google is currently one of the most widely used frameworks that come with built-in parallelism that enables usage of multi-threading to optimize learning. Keras is a high-level API that uses popular frameworks as the backend to run deep learning models. The model zoo provides the open-source implementation of the latest state-of-the-art neural networks in these frameworks. The hardware used in the training of the various methodologies discussed in the next section is as follows:

- GPUs—NVIDIA TitanX, NVIDIA Tesla K20c, and
- CPU—Intel i5-4570 CPU @ 3.20 GHz.

4 Deep Learning Architectures

Ever since the breakthrough deep learning model AlexNet won the ImagNet competition of 2012, many deep learning models have risen each bettering the previous best model. The inception [4] model was the winner of the 2014 ILSVRC 2014 image classification challenge extension of which the equally famous is Inception-v3 [5] network. The inception architecture shot into fame by making neural networks even deeper by stacking a series of inception modules. These modules consisted of parallel filters that operated on the input whose outputs were cascaded so that high-level information is not lost. The VGG-16 [6] on the other hand repeated 3×3 convolutions with three huge FC layers completing the structure. Both these architectures were designed to classify ImageNet data of 1000 classes, and trained on cluster of GPUs. Hence, fine-tuning and transfer learning use pretrained weights and modify the last few layers of these models, keeping in context the deep feature extraction by these models.

5 Classification Methodology

5.1 Training a Small CNN from Scratch

The first method constituted training a CNN from scratch using the medical data. The layer architecture is three convolutional layers with 3×3 kernels of numbers 32, 32, and 64, respectively (Fig. 1).

The features from the FC layer were classified using a regular sigmoid function into the two classes benign or malignant. The training of the CNN was done on the GPU and since the model is relatively shallow and the dataset small, the training is completed in an hour's time.

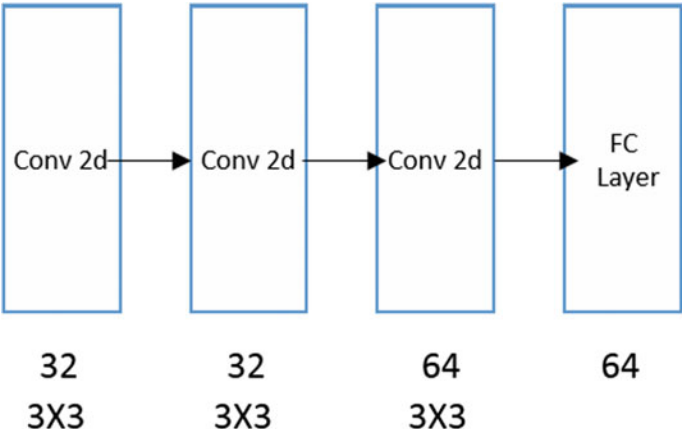


Fig. 1 CNN architecture which was trained from scratch

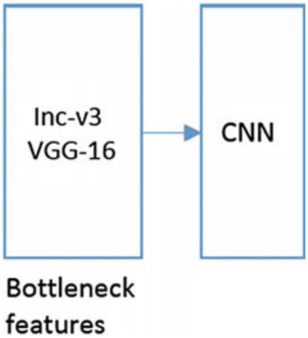


Fig. 2 Bottleneck features + CNN

5.2 Transfer Learning

Bottleneck features and the CNN: The bottleneck features from the VGG-16 and Inception-v3 were obtained and then trained on the CNN from the previous method. In VGG-16, the last three FC layers were discarded and the CNN model which we used in the first method was fed these features. The ImageNet pretrained weights were loaded into the models and after the forward pass of the image through the network bottleneck features was saved (Figs. 2 and 3).

Bottleneck features and SVM: The CNN was replaced with the popular linear classifier support vector machine which was fed the bottleneck features for classification. The simple default parameters of the SVM implementation provided by the scikit-learn were used in this method. Deep CNNs are known to be excellent feature extractors and using linear classifier to use these features proves to be an excellent way of tackling smaller datasets [7].

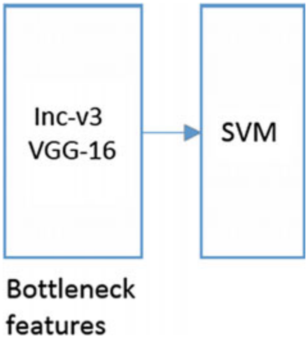


Fig. 3 Bottleneck features + SVM

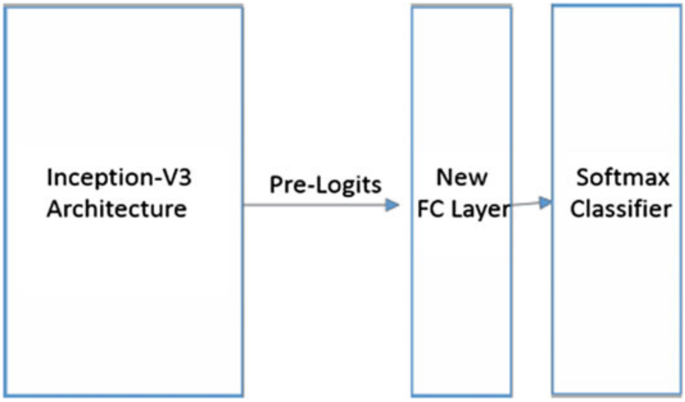


Fig. 4 Fine-tuning Inception v-3

5.3 Fine-Tuning Inception-v3 and VGG-16

The Inception-v3 model was imported with the help of tf-slim high-level API provided by TensorFlow. With the help of checkpoints provided for each of the models, pretrained models could be availed and fine-tuned. The Inception-v3 net provided in the slim API returns the list "end points" and "logits" which can be fed to a classifier to predict the class. We obtained the end points["pre-logits"] which is a layer prior to the last layer in the architecture and customized the FC layer, to give output as a binary classifier. Softmax classifier was used for the classification (Figs. 4 and 5).

The last three FC layers of the VGG-16 which contribute to huge computations were discarded, and new FC layer was attached after the "pool-5" layer. For fine-tuning, all layers above the conv 5_2 were frozen. So essentially the last three layers (excluding the FC layers) and the new custom FC layer were trained in this approach.

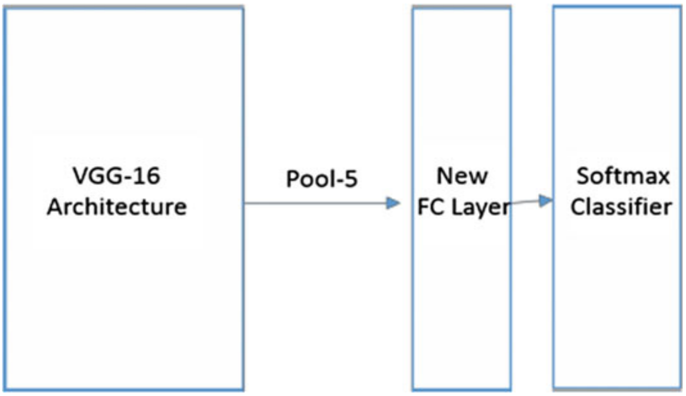


Fig. 5 Fine-tuning VGG-16

Table 1 Summary of classification results on the combined dataset

Sl No.	Model	Accuracy	Sensitivity	Specificity
1	CNN from scratch	0.82	0.82	0.89
2	VGG-16 + CNN	0.79	0.84	0.74
3	Inception-v3 + CNN	0.93	1.0	0.87
4	VGG-16 + SVM	0.94	1.0	0.88
5	Inception-v3 + SVM	0.89	0.88	0.89
6	VGG-16 fine-tuning	0.89	0.88	0.9
7	Inception-v3 fine-tuning	0.79	0.77	0.8

6 Results

The metrics used for evaluating the aforementioned methods are accuracy, i.e., ratio of the total number of correct predictions to the total number of images predicted, sensitivity, which gave an indication of true positive rate and specificity for true negative rate. The classification was done on both the datasets separately and on the combined dataset. It was observed that the first public dataset gave high sensitivity and low specificity, while the second dataset gave just the opposite. This is due to the nature of the data, wherein the first public dataset consisted of biased data with number of cancerous samples on the higher side. The local dataset on the other hand had the bias toward normal samples. This problem was handled by combing the datasets and also data augmentation achieved by flipping, rotating, and adding noise to the existing images. The table below summarizes the results obtained on the combined dataset alone. This combined dataset consisted of 2525 training samples and 613 test samples. The metrics are tabulated as percentages (Table 1).

7 Discussion

The delicate balance between the various evaluation metrics is important for performance analysis of deep learning algorithm in biomedical imaging. Owing to the fact that the dataset size is small, fine-tuning the deep architectures resulted was bettered by the other approaches. The amalgamation of linear classifier like the SVM to the architecture (especially the VGG-16) turned out to be the best model with stable metrics. Linear classifiers are considered to be useful when the deep neural networks are trained on datasets which are very different from domain in question. In order to take the classifier to a useful deployment, we are currently exploring system-based GUI and mobile-based android application of TensorFlow. Also, multi-class classification of the various thyroid ailments is being explored along with data collection and ground truth establishment in this regard.

Acknowledgements We would like to express our gratitude to our founder chancellor Bhagwan Sri Sathya Sai Baba for His inspiration. We would like to thank Department of Physics and Department of Mathematics and computer science for their support and also Dr. Narsimhan and Dr. Trimurthy from the Sri Sathya Sai Mobile hospital for their valuable insights. Consent of all participants was taken.

References

1. Pedraza L, Vargas C, Narváez F, Durán O, Muñoz E, Romero E (2015) An open access thyroid ultrasound image database
2. Abadi M, Agarwal A, Barham P, Brevdo E, Chen Z, Citro C, Corrado GS, Davis A, Dean J, Devin M et al (2017) Tensorflow: large-scale machine learning on heterogeneous distributed systems. arXiv preprint arXiv:1603.04467
3. Chollet F Keras. Github repository, https://github.com/fchollet/keras
4. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A (2015) Going deeper with convolutions. In: The IEEE conference on computer vision and pattern recognition (CVPR)
5. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z (2017) Rethinking the inception architecture for computer vision. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 2818–2826
6. Simonyan K, ZissermanVery A (2015) Deep convolutional networks for large-scale image recognition, arXiv:1409.1556v6
7. Sharif Razavian A, Azizpour H, Sullivan J, Carlsson S (2014) CNN features off-the-shelf: an astounding baseline for recognition, arXiv:1403.6382
8. Chi J (2017) Thyroid nodule classification in ultrasound images by fine-tuning deep convolutional neural network. J Digit Imaging 30:477–486
9. LeCun Y, Bengio Y, Hinton G (2015) Deep learning. Nature 521(7553):436–444
10. Géron A (2017) Hands-on machine learning with Scikit-learn and tensorflow. O'Reilly Media, Inc., 1005 Gravenstein Highway North, Sebastopol
11. Codebox.: Image_augmentor, https://github.com/codebox/image_augmentor
12. Lee J-G, Jun S, Cho Y-W, Lee H, Kim GB, Seo JB, Kim N (2017) Deep learning in medical imaging: general overview. Korean J Radiol 18(4):570–584

Medical Applications of Additive Manufacturing

A. Manmadhachary, Santosh Kumar Malyala and Adityamohan Alwala

Abstract Additive manufacturing (AM) is also known as "3D printing." AM creates the physical model by construction of succeeding layers using the input material. Each succeeding layer is attached to the preceding layer to form the physical AM model. Medical industry requires the error-free exact anatomy of the patient for diagnosis, surgical planning, surgical guides, implants, etc. Since every patient has unique anatomy, AM suits as best fit for medical industry. AM medical model provides advantage of customizing each model according to the patient's specific requirement, which is not so easy in conventional way. The current work aims to explain the importance of AM medical model in the medical industry for various scenarios.

1 Introduction

Chunk Hull has got the thought of AM in the year 1986 and named it as stereolithography apparatus (SLA). The working principle behind the AM is same as inkjet printer. To generate a 3D or physical model, the input data is initially generated into form of layers using a preprocessing software [1]. AM models are to be fabricated using computer-aided design (CAD), digital imaging and communications in medicine (DICOM), and any reverse engineering (RE) data. AM provides the flexibility to change or modify the design on the models easily at any stage,

A. Manmadhachary (✉)
Department of Mechanical Engineering, Faculty of Science and Technology,
ICFAI University, Hyderabad, Telangana, India
e-mail: manmadhachary@yahoo.co.in

S. K. Malyala
Department of Mechanical Engineering, National Institute of Technology,
Warangal, Telangana, India

A. Alwala
Department of Oral and Maxillofacial Surgery, Panineeya Institute of Dental Science
and Research Centre, Hyderabad, Telangana, India

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_152

without any extra efforts or wastage in time (till the fabrication stage). One of the most benefited industries using AM is medical [2]. AM provides an advantage to build multiple custom models in single batch. In a batch, each and every part can be customized as per the demand, which exactly needed in medical industry. As per recent advancements, AM has entered into the mass customization of implants for medical industry. Another major advantage with AM is any complex part that can be fabricated very easily with better accuracy. Medical industry is related to the life of the people where quality of part is most important [3]. This industry is responsible for curing of patient from suffering in a significant way to improve the quality of patient lives. Currently, AM medical models are using in dental, cosmetic surgery, hip and knee joints, bone fractures, tumor resection, etc. [4]. Apart from complex cases, the AM is also widely used in dental industry like casting models, bridges, crowns, and abutments. Most commonly used materials in medical industry are stainless steel, cobalt-chrome, titanium, and its alloys. Even though currently lots of materials are available in AM, majorly already certified milled or cast materials are easily available options for immediate use [5]. Compared to the conventional machining, AM provides superior flexibility and easy to use [6, 7]. This AM allows to reduce the number of sub-assemblies and number of parts in a product. The current chapter explains the importance of AM technology in anatomy visualization, preplanning surgery, implant manufacturing and preparation of surgical guides.

2 Procedure for Manufacturing of AM Medical Model

The development of surgical procedures using AM is shown in Fig. 1 [1]. The patient data acquired through the CT or MRI in form of DICOM will be processed through medical processing software like 3-Matic, MIMICS, 3D Doctor, 3D Slicer, etc. The data at acquisition stage is stored in the 2D format with equal thickness, which can be also called as 2½ D data. This 2½ D data can easily be converted into 3D data using medical processing software. From the complete acquired 3D data, required region of interest is separated and the 3D CAD data is generated. In process of separating required region of interest, the color, texture, and properties can be changed. The required information like bony or soft tissue is separated using Hounsfield units (HU) values from DICOM data. The required region of interest CAD model is converted into stereolithography (STL) model, which is globally accepted by all the AM machines. The STL format is surface representation of 3D CAD model in the form of triangular facets. This STL data is used as input for the preprocessing software to generate AM machine-specific code to fabricate the physical model. The AM machine reads data and lays downs or adds successive layers of liquid, powder, sheet material, or other, in a layer-upon-layer fashion to fabricate a 3D object. Few clinical cases that are solved using AM technique are explained in the later section.

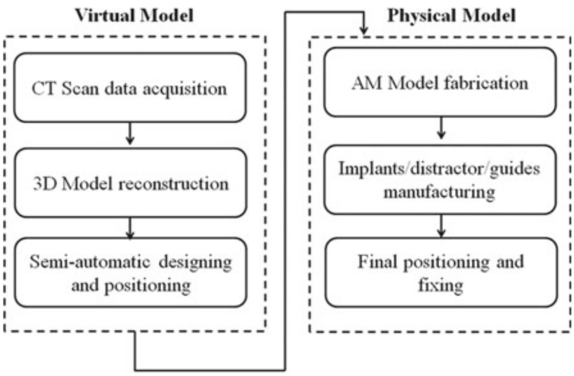


Fig. 1 Development of surgical procedures using AM medical model

3 Medical Applications

In the current chapter, all clinical case studies were conducted and patient's DICOM images procured from department of oral and maxillofacial surgery, Panineeya Institute of Dental Science & Research Center, Hyderabad.

3.1 AM in Visualization

AM model provides better understanding and visualization for the complex structures of anatomy, which is difficult to analyze using conventional 2D information. The simplest implementation of AM anatomical models into "bone resorption" is explained with a case study. The patient CBCT images were used to develop CAD model further AM model is developed. This AM model represents the patient anatomical perspective, resolution, and scale. This AM model is used to analyze the bone width in different places of maxilla and mandible, position of inferior alveolar nerve (IAN), estimation of maxilla sinus, etc. Based on this AM model, the surgeon plans implant fixing position and explain to the patient difficulties of surgery, as is shown in Fig. 2.

The below case study explains the importance of AM medical model in visualization of human anatomy, where patient (29-year male) with an accident. The complexity of this case is that patient had already met with accident in the same region, which increased complexity of the surgery. The previous surgeries metal plates present in the same region create artifacts in the CT. Due to metal artifacts, radiologist and doctor cannot identify exact location of damaged area. These CT images are processed in medical software using filters to remove artifacts and develop CAD model, as is shown in Fig. 3a. AM model was developed using CAD model; AM model is shown in Fig. 3b. Based on this AM model, surgeon identified the exact location, size, and

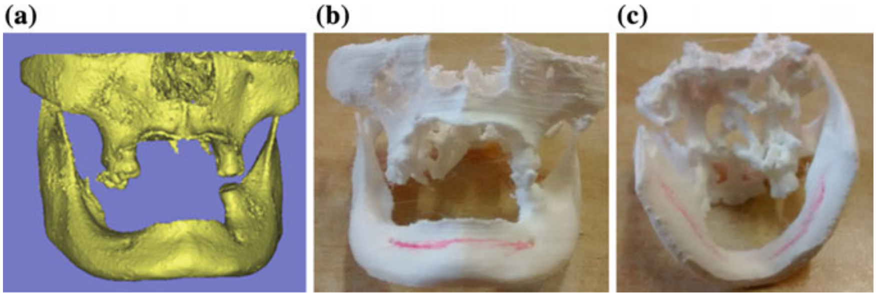


Fig. 2 **a** CAD model, **b** Front side AM model, and **c** Backside AM model

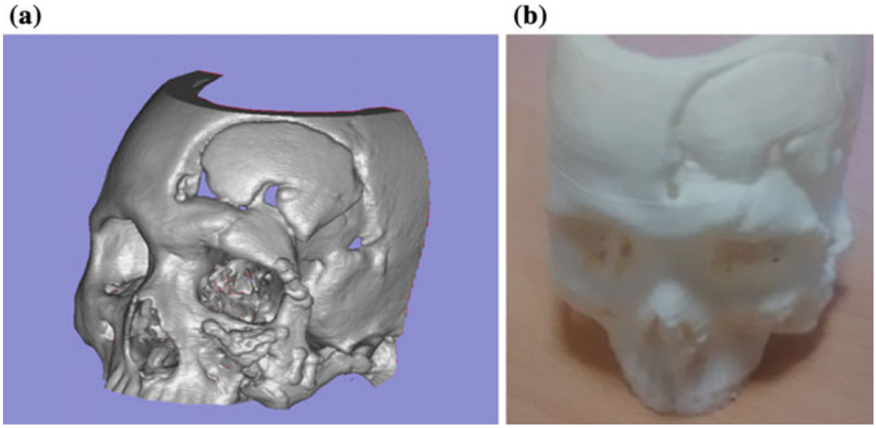


Fig. 3 **a** CAD model and **b** AM models of patient anatomy

shape of damaged area. Also, this AM model was used to bend the reconstruction plates for actual surgery, as it reduced the time for reconstruction of the frontal area.

3.2 AM in Preplanning Surgery

AM models represent exact dimensions of anatomical models of the human anatomy. AM medical models are used by surgeons to diagnose the complex structures anatomy. These AM models can be used as trial models by surgeons in case of complex surgeries. In this chapter, mandibular reconstruction and pan facial trauma surgeries were conducted using AM models. These are explained in the following sections.

This study deals with a 20-year-old male, who had a tumor in right side of mandible, as shown in Fig. 4a. He was reported with swelling and loose teeth in the right mandible. Mandibular re-establishment is challenge for the surgeon wish-

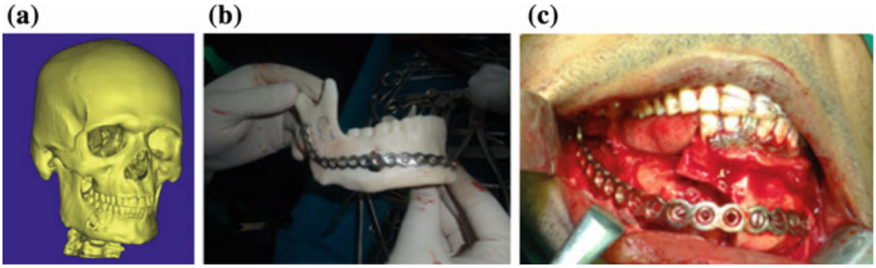


Fig. 4 **a** Tumor in mandible, **b** Bending of plate on AM model, **c** Resection of tumor and fixing of plate

ing to reconstruct its unique geometry. Reconstruction can be attained with titanium plates followed by autogenous bone grafting. Incorporation of the bone graft into the mandible provides continuity and strength required for proper esthetics. Dental implant rehabilitation at a later stage. Valuable time in the operating room is consumed in plate contouring to reconstruct the mandible. The tumor mandible CAD model was developed by using DICOM images, as shown in Fig. 4a. The CAD model is used to develop physical AM model. Mandible AM model was manufactured before the resection of tumor mandibular, which is used to prepare bridging plates before the first stage of reconstruction, as shown in Fig. 4b. Based on this AM model, the expected resection margins were designated, and a KLS Martin reconstruction plate was bent, adapted to the shape of the mandible as shown in Fig. 4c. The time needed to pre-bend the plates was minimized. The tumor was analyzed and removed based on AM model, as shown in Fig. 4c.

The current case is pan facial trauma which was treated by AM medical model. The AM medical model was used for presurgical planning [8]. Reconstruction of face can be attained with fixing of titanium plates followed by autogenous bone grafting. The patient maxilla and mandible CAD model was developed by using DICOM images, as shown in Fig. 5a. From the CAD model, we can observe the fractures of the patient maxilla, mandible and the nasal regions. The original CAD model is used to split the various fractured regions by using MIMICS software, as shown in Fig. 5b. The fractured regions of CAD models were used to develop the AM models. The various fractured regions are manufactured with different color filaments in AM process. These AM parts are used to bend and fixing of plates in fractured regions. The reconstruction plate was bent, adapted to the shape of the maxilla, mandible, and nasal regions, as shown in Fig. 5c. The time needed to pre-bend, location, and fixing the plates was minimized. These presurgical procedures and precautions are used to surgery in operation theatre, as shown in Fig. 5d and e.

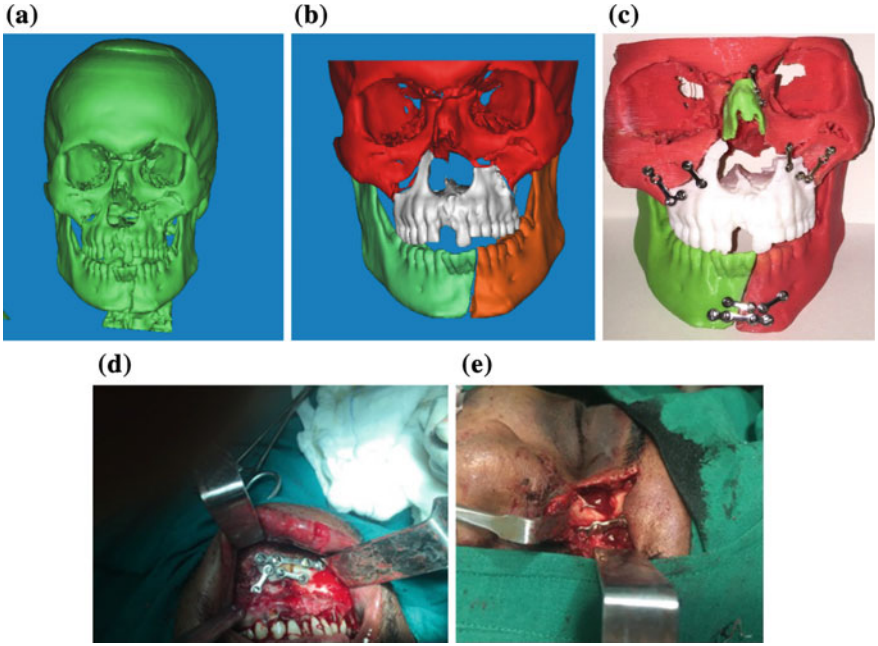


Fig. 5 **a** Patient CAD model, **b** Fractures in CAD model, **c** preplanning surgery on AM model, **d** fixing of plates at mandible region, **e** fixing of plates at maxilla region

3.3 AM in an Implant Design

The implant design and manufacturing with using patient-specific AM models are one of the trend in medical applications. These implants were exactly fixed to anatomical models in the human body [9]. In this chapter, design and manufacturing of distractors for cleft and tumor mandibular were developed with using AM models. These are explained in the following sections.

The current case study is regarding a 35-year male patient, the patient presented with a defect in maxillary bone which has been diagnosed as cleft alveolus based on history and clinical evaluation. It is associated with regurgitation of food and fluids into nose on consumption. On examination of CAD model, the size of the defect has been recorded as 16.6 mm based on measurement of cleft diameter in CAD model; these are shown in Fig. 6a, b at front and back side views, respectively. This CAD model was used to build a physical AM. This AM model is used for the preparation of distractor. Here, the bending, length and fixing of distractor were evaluated by using AM model. The distractor movement after bone extraction from cleft part is also planned with using AM model. The various movements of distractor are shown in Fig. 6c–e. These same preplanning surgical procedures are used at the time of actual surgery, as shown in Fig. 6f, g.

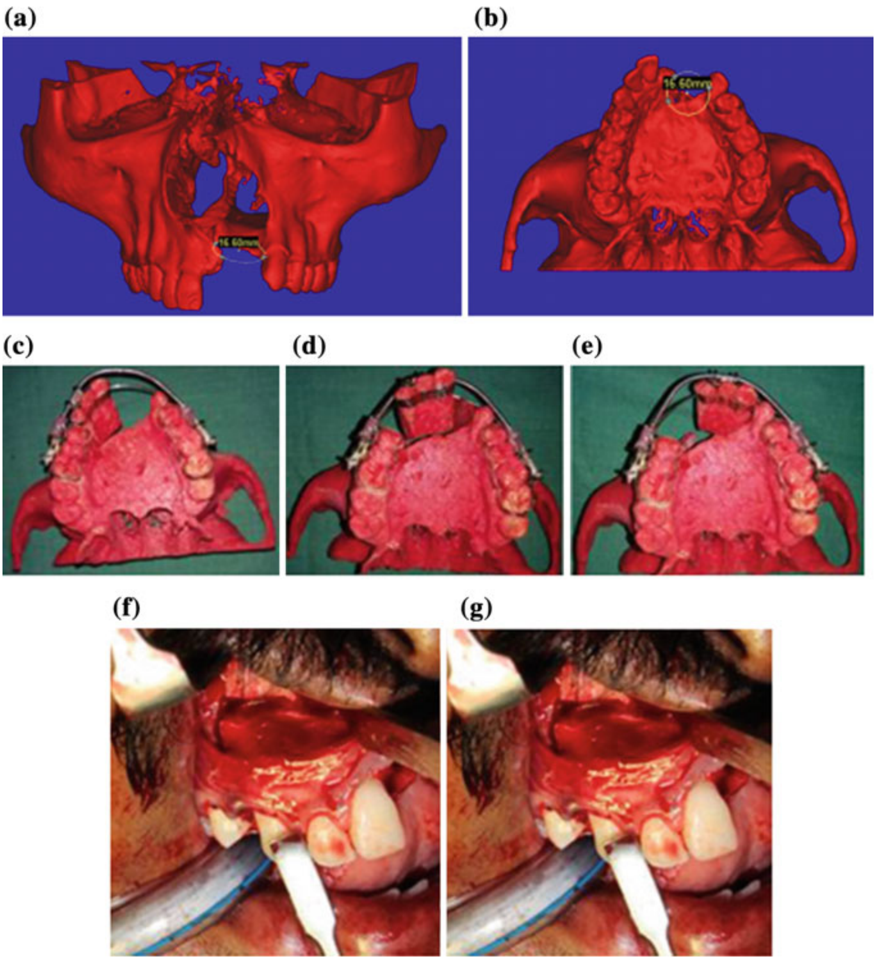


Fig. 6 Cleft in **a** front view, **b** bottom view, **c**–**e** surgical procedure for bone movement, **f**, **g** fixing of patient-specific distractor

The current case study is described to reconstruct a mandibular continuity defect using a technique of bifocal distraction osteogenesis, causing minimal morbidity to the patient [10]. The CT images of patient were reconstructed a CAD model, incompletely reconstructed mandibular defect, as shown in Fig. 7a. Further, AM model was developed; this defect was reconstructed using a custom-made distractor, as shown in Fig. 7b. This distractor was developed based on patient-specific dimensions such as length, width, and angle of mandible. These measurements are measured and aligned based on AM model. A mandibular segmental defect was reconstructed using the bifocal distraction technique, with the residual defect being reconstructed using an anterior iliac block graft, hence avoiding a microvascular reconstructive procedure.

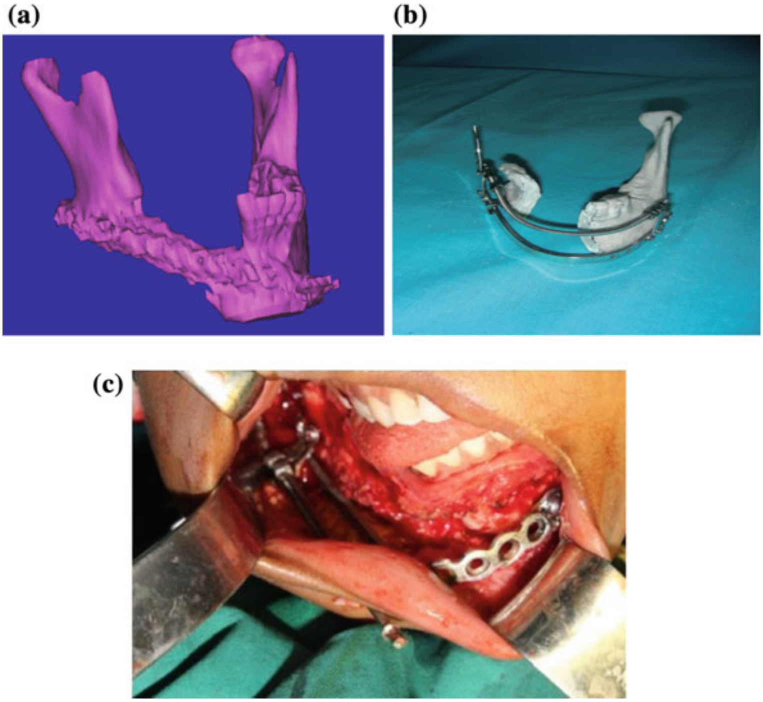


Fig. 7 **a** Mandibular defect in CAD model, **b** Patient-specific distractor and preplanning, **c** Fixing of distractor

The procedure was planned to completely avoid a bone grafting or a microvascular procedure to reconstruct the lost mandible. However, complete regeneration was reconstructed with using patient-specific distractor. Thus, patient obtained a satisfactory facial symmetry and functionally continuous mandibular bone for daily function of speech.

3.4 AM in Design and Manufacturing of Surgical Guides

AM is changing design and manufacturing of surgical guide in dentistry. Till today, the processes for dental restorations, denchers, bridges, and implants, have been limited to lost-wax technology, a method that has been used for the past 100 years. The denchers made with lost-wax method start with a gypsum model. Based on this model, the surgical guides were preparing for dental surgery. The new technologies are coming on-stream to answer that need. The FDM AM process and the material PLA are used to produce dental models [11]. The dental plastic model (mandible) models can be manufactured within 4 h depending on the size. What is more, the

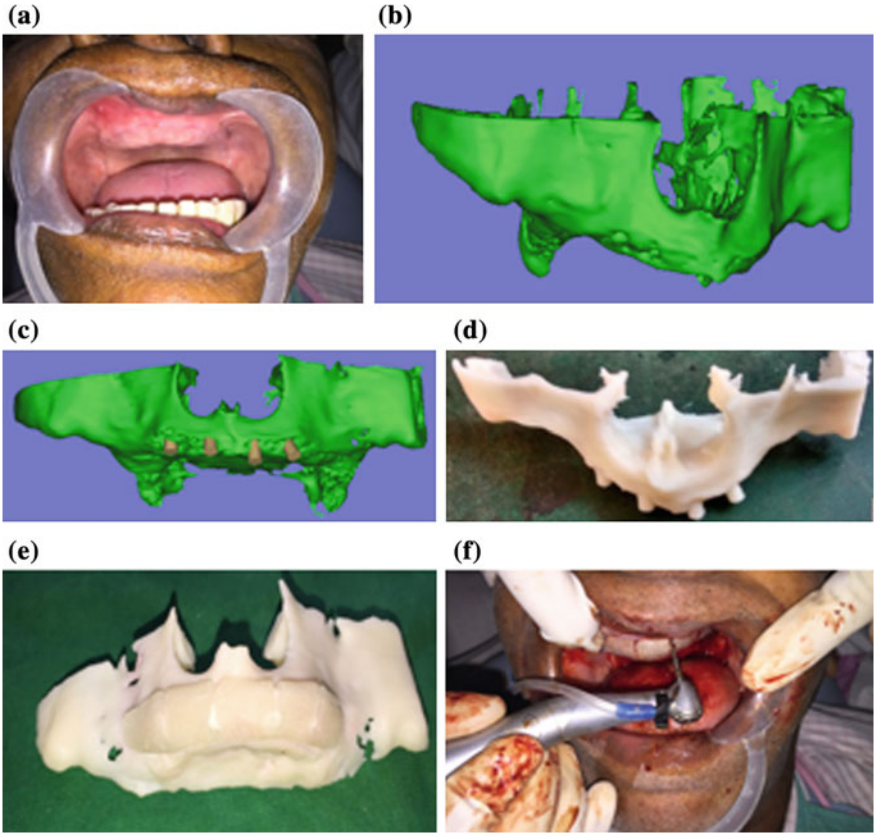


Fig. 8 **a** Patient maxilla, **b** CAD model, **c** Implants are fixed on CAD model, **d** AM model, **e** Preplanning surgery, **f** Actual surgery

plastic models are very hard-wearing, an important requirement if the crowns and tooth stumps have to be fitted and removed several times. The current case study is used for fixing of implants to set artificial teeth in maxilla, the patient maxilla as shown in Fig. 8a. 3D CAD was generated by MIMICS software from DICOM images of the patient, which is shown in Fig. 8b. Further, this CAD model was used for fixing of implants, with accurate dimensions and angels virtually, as shown in Fig. 8c. This implants designed CAD model was used to manufacture AM model, as shown in Fig. 8d. This AM model along with patient-specific implant gives surgeons to anticipate the steps in actual surgery. This AM model is used for design and manufacturing of surgical guide to fixing of implants, as shown in Fig. 8e. These AM medical model and surgical guide are using surgeons to perform a surgery for implantation with accurate dimensions and angles for the patient and also avoid the damage of superior alveolar nerve, as shown in Fig. 8f.

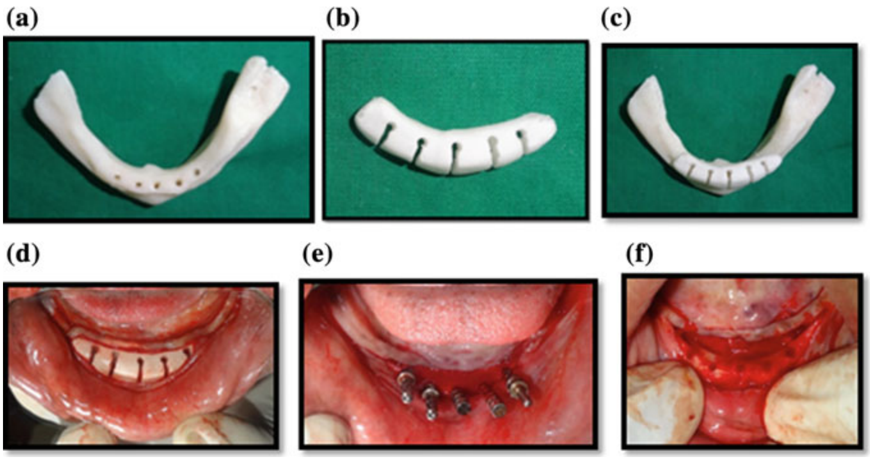


Fig. 9 **a** AM model, **b** Customized surgical drilling guide (CSDG), **c** CSDG on AM model, **d** CSDG for drilling holes on mandible, **e** Implants placement in mandible, **f** Post-operative surgery

A customized surgical drilling guide (CSDG) has proven to be beneficial to dentistry, for accurate placement of osseointegrated implants in partially mandible in flapless surgery [12]. The FDM AM process and the material PLA are used to produce a dental model, as shown in Fig. 9a. As per the need, these CSDGs have made bone or mucosa or gingival supported for the exact placement of implants, which is shown in Fig. 9b. The FDM model and CSDG model were used for preplanning of implantation for exact alignments, as shown in Fig. 9c. The surgeon's use of the CSDG allows the orthodontist and the surgeon to exact placement of implants at correct and safe position with firm and stable placement, as is shown in Fig. 9d, e and f shows placement of implants in the mandible.

4 Conclusions

The AM technology provided a precise, fast, and cheap anatomical reconstruction, which aids in shortened operation time (and therefore decreased exposure time to general anesthesia, decreased blood loss, and shorter wound exposure time) and easier surgical procedure. The application of AM in visualization was allowed to evaluate the patient anatomy conditions before and after surgery. The preplanning surgical models help medical specialists to diagnose ailments and plan for surgery. These AM models are widely used in complex surgery to ease the planning process. The implant design and manufacturing for reconstruction of the bony defect surgeries are conducted with AM medical models. Treatment planning and mock surgery were performed on the AM medical model, which reduced about 25–30% of total surgery time, thereby decreasing the complications.

Acknowledgements The authors would like to special thanks for their support, Dr. N. V. S. Shekar Reddy and Dr. U. D. Arvind, Department of Oral and Maxillofacial Surgery, Panineeya Institute of Dental Science and Research Center, Hyderabad, India for providing DICOM images and conducting surgeries in their hospital. No ethical approval was needed for these studies and all the local guidelines were followed.

References

1. Mamoru MJC, Paulo B, Dirk F, Albert JS, Kamlakar R, Naohiko S (2013) Bio-manufacturing. CIRP Ann Manuf Technol 62:585–606
2. Wohler TT (2013) Additive manufacturing and 3D printing state of the industry, Annual worldwide progress report. Wohlers Associates, Colorado, USA
3. Webb PA (2000) A review of rapid prototyping (RP) techniques in the medical and biomedical sector. J Med Eng Technol 24:149–53
4. Parthasarathy J (2014) 3D modeling, custom implants and its future perspectives incraniofacial surgery. Ann Maxillofac Surg 4:9–18
5. Farias TP, Dias FL, Sousa BA, Galvao MS, Bispo D, Past AC (2013) Prototyping: major advance in surgical planning and customizing prostheses in patients with bone tumors of the Head and neck. Int J Clin Med 4:1–7
6. Manmadhachary A, Ravi Kumar Y, Krishnanand L (2018) Finding of correction factor & dimensional error in bio-AM Model By FDM Technique. J Inst Eng (India): Ser C 99 https://doi.org/10.1007/s40032-016-0294-1
7. Manmadhachary A, Kumar MS, Ravi Kumar Y, Haranadha Reddy M, Adityamohan A (2017) Design & manufacturing of Implant for reconstructive surgery: a Case Study. KnE Eng 2:143–149
8. Alwala AM, Malyala SK, Chittaluri LR, Vasamsetty P (2016) Surgical planning in pan facial trauma using additive manufacturing medical model-a case study. Journal of Surgery Jurnalul de chirurgie. 12:125–128
9. Santosh Kumar Malyala Y, Kumar R, Alwala AM, Manmadhachary A (2017) 3D printed medical model to resolve cleft alveolus defect: a case study. KnE Eng 2:8–14
10. Alwalaa AM, Arvind UD, Malyala SK, Vasamsetty P (2018) Customization of patient specific distraction device using additive manufacture technology. Mater Today: Proc 5:4134–4137
11. Manmadhachary A, Ravi Kumar Y, Krishnanand L (2016) Improve the accuracy, surface smoothing and material adaption in STL file for RP medical models. J Manuf Process 21:46–55
12. Malyala SK, Ravi Kumar Y, Kankanala L, Vasamsetty P, Alwala A (2018) Assessment and treatment planning in maxillofacial surgery by using additive manufacturing technology. Mater Today: Proc 5:4162–4166

A Study on Comparative Analysis of Automated and Semiautomated Segmentation Techniques on Knee Osteoarthritis X-Ray Radiographs

Karthiga Nagaraj and Vijay Jeyakumar

Abstract Arthritis is a most common disease in the worldwide population targeting knee, neck, hand, hip, and almost all the joints of the human body. It is a frequently noticed problem in elder people, especially women. The severity of the disease is analyzed using the older KL grading system. Traditionally, the detection of various grades of OA (osteoarthritis) is interpreted by just a visual examination. A traditional modality, X-ray images are considered as the data for the project. The images are segmented using different segmentation techniques to extract the articular cartilage as region of interest. From the literature, eight different segmentation techniques were identified out of which seven are automated and one is semiautomated. By implementing those techniques and evaluating their performance, it is inferred that block-based segmentation, center rectangle segmentation, and the semiautomated seed point selection segmentation performs well and provides sensitivity, positive prediction value and dice Sorenson's coefficient of 100%, respectively, and specificity of 0%.

1 Introduction

Arthritis is a common disease worldwide. Osteoarthritis is the most common type of arthritis. It is also called as disk degeneration, which resembles the degeneration of the articular cartilage sandwiched between femur and tibia [18]. As the name signifies, osteoarthritis means inflammation in joints of bones. Generally, the degree of movement of any joint is from 0 to 120°, but osteoarthritis-affected individuals' faces restriction and pain during mobility (Fig. 1).

K. Nagaraj · V. Jeyakumar (✉)
Department of Biomedical Engineering, SSN College of Engineering, Chennai, India
e-mail: vijayj@ssn.edu.in

K. Nagaraj
e-mail: karthiga16455005@bme.ssn.edu.in

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_153

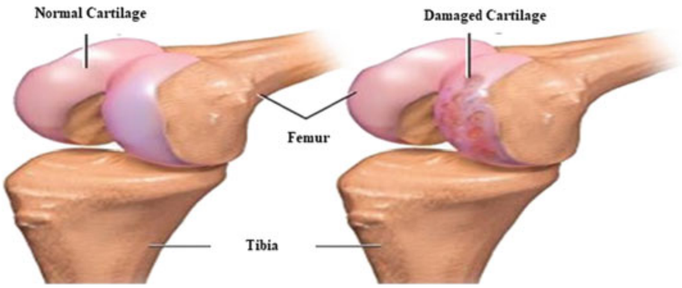
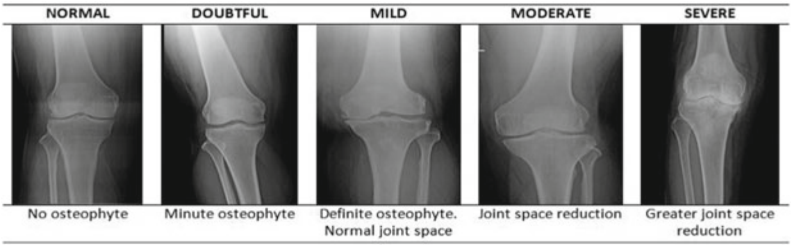


Fig. 1 Normal and OA affected knee (*Source* http://pennstatehershey.adam.com)

Table 1 Different grades of OA based on Kellgren–Lawrence system (*Source* tdmu.edu.ua)



The knee disks (cartilages) are the fibrous flexible bones and they are flexible due to its high water content. These cartilages are present in all the joints and they serve as the cushion for the joints to prevent wear and tear of joints, which indirectly inhibit bone friction [6]. But, due to few other factors like, aging, obesity, over load bearing works, hereditary, hormonal imbalance, gender, trauma, etc., the cushion between the joints gets degenerated [14] and this leads to severe pain [1]. Kellgren and Lawrence (KL) system classifies osteoarthritis into five distinct grades based on disease severity. This system was proposed by Kellgren and team in 1957 and was accepted by WHO (World Health Orgaization) in 1961. The grading ranges from 0 to 4. Grade 0 means there is no characteristic symptom of OA is observed. Grade 1 means significantly doubtful symptoms of OA observed. Grade 2 means mild disk narrowing has been observed. Grade 3 means moderate narrowing of the cartilage is observed. Grade 4 means the disease is at its severe stage. Out of many systems of classification [3], this is found to be used widely by the doctors, especially in India (Table 1).

2 Materials and Methods

The real-time X-ray dataset of 300 patients in the age group of 25–80 were acquired at Manisundaram Medical Mission Hospitals, Vellore under the assistance of

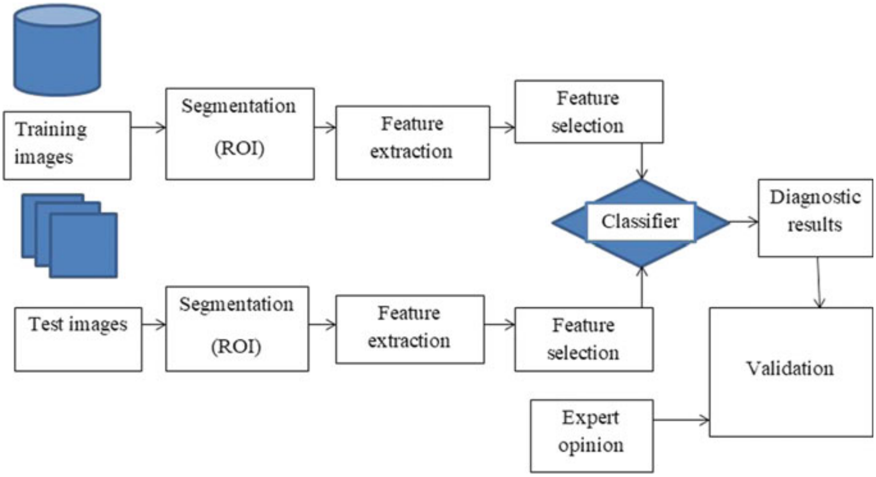


Fig. 2 Workflow of the proposed method

Dr. Manivannun K, Mr. Selva Prakash S, and Ms. Suganya S. The entire framework was performed with MATLAB 2017a (Fig. 2).

The images were acquired using X-ray equipment Wipro GE-DX 300 and preprocessed using the software Imageworks. The sample set of 25 images were taken and different kinds of segmentation techniques were applied on to them.

The proposed work used 90% of the acquired images as the training data and 10% as the test data. The images are initially segmented to impart the region of interest. Then, the segmented images proceeded with the feature extraction. From the extracted features, important features are to be selected and those features are used in training a conventional neural network [8, 10]. The obtained results are to be validated by a medical practitioner.

In this work, we have worked on segmentation using automated and semiautomated segmentation techniques. The complication in segmentation of the articular cartilage is that its gray intensity lies amidst in the gray scale and the X-ray contains various intensities of gray. Background removal fails as the cartilage and the unwanted image portion (the surrounding muscles) falls under similar gray intensities [12].

3 Segmentation

Image segmentation is a process in which information is extracted from an image. Segmentation is done on the preprocessed image to segment bones from the X-ray image, thereby to highlight the cartilage. Different segmentation techniques were [15–17]applied to find the better option [5].

The eight opted segmentation techniques are:
Automated segmentation techniques:

- Block-Based Segmentation
- Center Rectangle Segmentation
- Fuzzy *C*-Means Segmentation [2]
- *K*-Means Clustering [4]
- Marker-Based Watershed Segmentation
- Morphological Operations Based Segmentation
- Region-Based Active Contour Segmentation.

Semiautomated segmentation technique:

- Seed Point Selection Segmentation.

Each segmentation method is further described in the following sections.

3.1 Automated Segmentation

3.1.1 Block Dıstance Measure Segmentation

This method is a simple technique where the image is being segmented in a lossless way [13]. Initially, the image is divided into nine blocks. Out of this nine blocks, the region of interest (RoI), cartilage is extracted from the blocks 4, 5 and 6.

3.1.2 Center Rectangle Segmentation

This technique is used to mark the region other than the region of interest. The image is resized initially to avoid memory error. Then the axes of the image from both X and Y coordinates are computed. Image parameters are then defined according to the size of the region of interest. It unmarks the cartilage region and leaves the other parts of the image unmarked [20].

3.1.3 Fuzzy C-Means Segmentation

The Fuzzy *C*-means algorithm is an unsupervised fuzzy clustering algorithm. Clustering or cluster analysis involves assigning data points to clusters such that items in the same cluster are as similar as possible, while items belonging to different clusters are as dissimilar as possible. Let $X = (x_1, x_2,\ldots,x_N)$ be an image with N pixels

$$J = \sum_{j=1}^{N} \sum_{i=1}^{c} u_{ij\,\|x_j - v_i\|^2} \tag{1}$$

where

x_i is a multispectral data,
u_{ij} is the membership function, and
v_i is the *i*th cluster center.

However, most soft clustering algorithms do generate a soft partition that also forms fuzzy partition. A type of soft clustering of special interest is one that ensures membership degree of point x in all clusters adding up to one.

3.1.4 *K*-Means Segmentation

K-means is a least-squares partitioning method that divides a collection of objects into *K* groups. The initial assignment of points to clusters can be done randomly. The potential point is

$$P_n = \sum_{i=1}^{n} \mathrm{e}^{\frac{-4x_n - x_i^2}{r_a^2}}, \tag{2}$$

where n is data points: $X = \{x_1, x_2, x_{3\ldots}\ x_n\}$.

The groups obtained are such that they are geometrically as compact as possible around their respective means. Using the set of feature images, a feature vector is constructed corresponding to each pixel. Each data pixel is

$$P_n = p_n - p_1 \sum_{i=1}^{n} \mathrm{e}^{\frac{-4x_n - x_i^2}{r_a^2}}, \tag{3}$$

where

p_1, p_n are first and n^{th} pixel, respectively

The *K*-means can then be used to segment the image into three clusters—corresponding to two scripts and background, respectively. For each additional script, one more cluster is added. Thus, even if a pixel is wrongly clustered, it can be corrected by looking at the neighboring pixels.

3.1.5 Marker-Based Watershed Segmentation

This algorithm is based on the image morphology [7]. It considers the topological gradient of the image. The boundaries of the image are segmented. The threshold of the image *T* is

$$T = \left\{ \max(i) | \sum_{j=0}^{i} H(j) \le 1 - \varepsilon \right\} \tag{4}$$

where $H(j)$ is the histogram of image j and ε is the appropriate percentage of pixels needs pixel suppression.

It marks the foreground and the background separately. A segmentation function is computed to segment the dark regions of an image. Then, the foreground markers of each image object are computed. Segmentation function is modified, henceforth it has only the minima of both foreground and background markers [11]. For the modified segmentation function, the watershed transform is performed.

3.1.6 Morphological Operations

Step 1: The image is subjected to median filter after being converted into gray image.
Step 2: The edge detection algorithm used is Sobel edge detection.
Step 3: Initially, the image is stretched and then it is dilated.
Step 4: The holes in the image after dilation is filled.
Step 5: The border of the image is cleared using the function.
Step 6: Finally, the outline of the image is estimated by marking its perimeter.

3.1.7 Region-Based Active Contour Segmentation

This method is applied to segmenting images to improve the bone estimation from the X-ray images. Active contours models are used to detect objects in a given image using the techniques of curve evolution [9]. It identifies individual segments in images with multiple segments and junctions. It uses Heaviside function for smoothing

$$H\varphi(x) = \begin{cases} 1, & \varphi(x) < -\varepsilon \\ 0, & \varphi(x) > \varepsilon \\ \frac{1}{2}\left\{1 + \frac{\varphi}{\varepsilon} + \frac{1}{\pi}\sin\frac{\pi\varphi(x)}{\varepsilon}\right\} \end{cases} \quad (5)$$

Mask is a binary image that specifies the initial state of the active contour. The boundaries of the object regions (white) in mask define the initial contour position used for contour evolution to segment the image. The output image is a binary image where the foreground is white (logical true) and the background is black (logical false).

3.2 *Semiautomated Segmentation*

3.2.1 Seed Point Selection Segmentation

In this technique, four random seed points from the image around the region of interest is selected. By selecting the seed points, we mean the boundaries of the ROI, i.e., the rows and columns of the corresponding boundaries of the ROI [19]. The buffer for the region of interest and non-ROI is developed, respectively. The marking of the interested region and the other parts of the image is done individually and the images are obtained.

4 Results and Discussion

The real-time X-ray dataset of 300 patients was acquired and it was analyzed with MATLAB 2017a. The sample set of 25 images were taken and different kinds of segmentation techniques were applied on to them. The results are as follows (Fig. 3):

Block-based aegmentation works well for all the grades where the blocks 4–6 combinedly contain the Region of Interest (ROI). The block 5 contains the highest portion of the cartilage (Fig. 4).

In center rectangle aegmentation, the center region which contains the cartilage is alone left unmasked, whereas the non-ROI regions are masked. This method goes well with the ground truth (Fig. 5).

Fuzzy *C*-means segmentation segments by forming clusters with similar pixels based on its gray level, distance, connectivity, and its intensity. It works good only with the denoised grade 1 OA radiographs (Fig. 6).

K-means segmentation method separates the foreground and background into two–three clusters based on their mean value amongst the pixel intensities in the image. For X-ray images, it produces an irrelevant clustering (Fig. 7).

Morphological operations method of segmentation includes almost all basic operations like denoising, opening, stretching, dilating, filling of holes, etc. Due to the

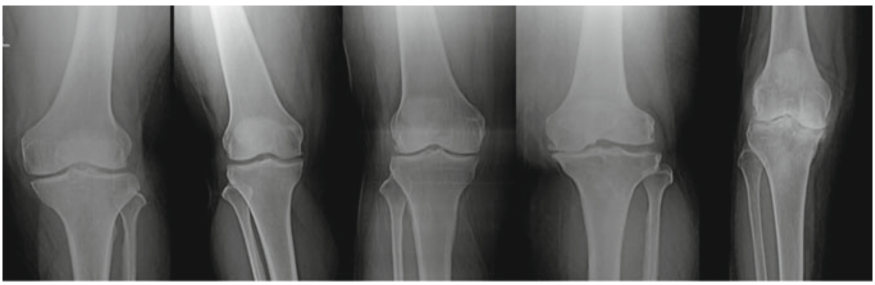


Fig. 3 Normal radiographs of grade 0–4 OA

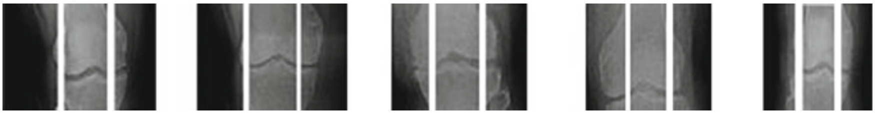


Fig. 4 Block-based segmentation for grade 0–4 OA radiographs

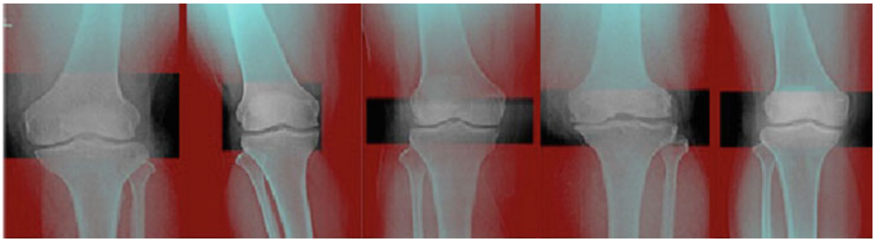


Fig. 5 Center rectangle segmentation for grade 0–4 OA radiographs

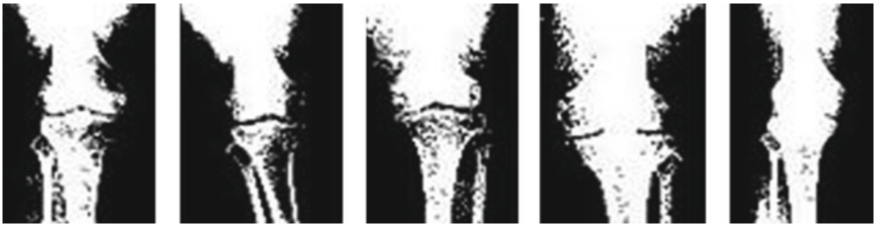


Fig. 6 Fuzzy *C*-means segmentation for grade 0–4 OA radiographs

Fig. 7 *K*-means segmentation for grade 0–4 OA radiographs

presence of noise in the real-time acquired images, this method is unfit for X-ray images (Fig. 8).

Region-based active contour segmentation technique segments the foreground of the image from the background of the image. It shows reasonable segments in radiographs of grade 0–2 (Fig. 9).

In marker-based watershed segmentation, the background and foreground are segmented considering the topological gradient of the image. The segments are

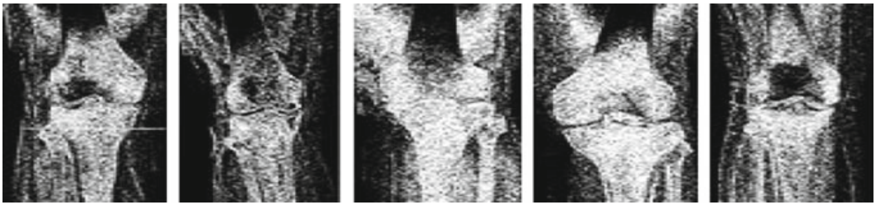


Fig. 8 Morphological operations for grade 0–4 OA radiographs

Fig. 9 Region-based active contour segmentation for grade 0–4 OA radiographs

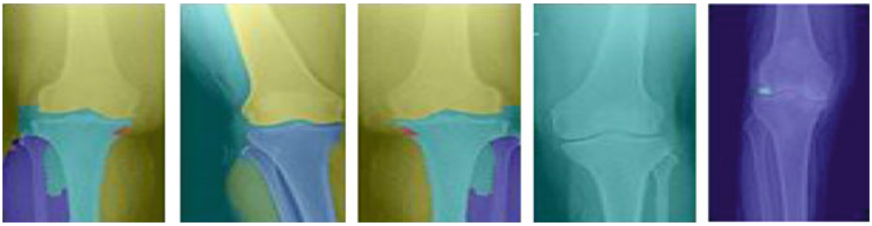


Fig. 10 Marker-based watershed segmentation for grade 0–4 OA radiographs

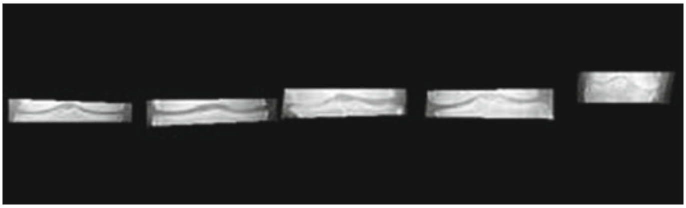


Fig. 11 Seed point selection segmentation for grade 0–4 OA radiographs

represented in the shades of RGB. This method works well for grade 1 and 2 images (Fig. 10).

Seed point selection segmentation method needs a manual selection of seed points, but it creates a defined boundary by tracing the rows and columns corresponding to the selected seed points. It goes well with radiographs of all the grades of OA (Fig. 11).

Table 2 Metrics for the performance of various segmentation techniques

Evaluation metrics (in percentage)	SE	SP	PPV	DSE
Ground truth	0	100	100	100
Segmentation techniques				
K-means clustering	0	4	4	4
Fuzzy *C*-means segmentation	100	75	50	5
Region-based active contour segmentation	100	0	60	10
Block-based segmentation	100	0	100	100
Marker-based watershed segmentation	100	40	40	7
Morphological operations-based segmentation	0	4	4	4
Center rectangle segmentation	100	0	100	100
Seed point selection segmentation	100	0	100	100

An evaluation metrics was computed for evaluating the effectiveness of various segmentation techniques and the parameters like sensitivity, specificity, positive prediction values, and the dice Sorenson's coefficients were used.

The confusion matrix is plotted for each and every segmentation method with its true positive, true negative, false positive, and false negative values estimated by an expert. With those confusion matrices, parameters like Sensitivity (SE), Specificity (SP), Positive Predictive Value (PPV), and Dice Sorensen Coefficient (DSC) are computed.

$$\mathrm{SE}\ (\%) = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} * 100 \tag{6}$$

$$\mathrm{SP}\ (\%) = \frac{\mathrm{TN}}{\mathrm{TN} + \mathrm{FP}} * 100 \tag{7}$$

$$\mathrm{PPV}\ (\%) = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} * 100 \tag{8}$$

$$\mathrm{DSC}\ (\%) = \frac{2.\mathrm{TP}}{2.\mathrm{TP} + \mathrm{FP} + \mathrm{FN}} * 100 \tag{9}$$

On the other hand, all the above mentioned parameters are also computed for the ground truth (free-hand cropped cartilage marked by an expert). The best working segmentation techniques that go hand-in-hand with the ground truth are selected for further processing (Table 2).

5 Conclusion and Future Works

It is hence concluded that center rectangle and Block-based segmentation which are automatic segmentation techniques and the seed point selection segmentation which is a semiautomated segmentation are found to perform well. The other methods were found uncompatible to comply with X-ray images. Hence, an optimal segmentation technique is inferred for segmentation of knee images for detection of osteoarthritis. The segmented X-ray images are compatible to be used for features extraction.

This work is to conclude with a good technique that works best in segmenting the knee X-ray images which will help in automatic classification based on KL grading. The future work is to deploy the fully automatic, convolutional neural network to classify the test images based on the KL grading system.

Acknowledgements The images used for this study were obtained from Manisundaram Medical Mission Hospitals, Vellore under the supervision of Dr. Manivannun K., M.S (Ortho.), Mr. Selva Prakash S., Dip. in X-ray Technology, and Ms. Suganya S B.Sc MSW, Counselor. We duly state that my data collection does not involve patient's interference and invasive protocol.

References

1. Boniatis I, Costaridou L, Cavouras D, Kalatzis I, Panagiotopoulos E, Panayiotakis G (2006) Osteoarthritis severity of the hip by computer-aided grading of radiographic images. Med Biol Eng Compu 44(9):793
2. Chuang KS, Tzeng HL, Chen S, Wu J, Chen TJ (2006) Fuzzy C-Means clustering with spatial information for image segmentation. Comput Med Imaging Graph 30(1):9–15
3. Deokar DD, Patil CG (2015) Effective feature extraction based automatic knee osteoarthritis detection and classification using neural network. Int J Eng Techn 1(3)
4. Dhanachandra N, Manglem K, Chanu YJ (2015) Image segmentation using K-Means clustering algorithm and subtractive clustering algorithm. Procedia Comput Sci 54(2015):764–771
5. Gan HS, Sayuti KA (2017) Comparison of improved semi-automated segmentation technique with manual segmentation: data from the osteoarthritis initiative
6. Gornale SS, Patravali PU Manza RR (2016) Detection of osteoarthritis using knee X-ray image analyses: a machine vision based approach. Int J Comput Appl 145(1)
7. Hill PR, Canagarajah CN, Bull DR (2003) Image segmentation using a texture gradient based watershed transform. IEEE Trans Image Process 12(12):1618–1633
8. Kubkaddi S, Ravikumar KM (2017) Early detection of knee osteoarthritis using SVM classifier. IJSEAT 5(3):259–262
9. Lankton S, Tannenbaum A (2008) Localizing region-based active contours. IEEE Trans Image Process 17(11):2029–2039
10. LeCun Y, Bengio Y, Hinton G (2015) Deep learning. Nature 521(7553):436–444
11. Lu S, Wang S, Zhang Y (2017) A note on the marker-based watershed method for X-ray image segmentation. Comput Methods Programs Biomed 141:1–2
12. Minciullo L, Cootes T (2016) Fully automated shape analysis for detection of Osteoarthritis from lateral knee radiographs. In: 2016 23rd international conference on pattern recognition (ICPR). IEEE, pp 3787–3791
13. Navale DI, Hegadi RS, Mendgudli N (2015) Block based texture analysis approach for knee osteoarthritis identification using SVM. In: 2015 IEEE international WIE conference on electrical and computer engineering (WIECON-ECE). IEEE, pp 338–341

14. Øiestad BE, Juhl CB, Eitzen I, Thorlund JB (2015) Knee extensor muscle weakness is a risk factor for development of knee osteoarthritis. A systematic review and meta-analysis. Osteoarthritis Cartilage 23(2):171–177
15. Rahman MA, Liu S, Lin S, Wong C, Jiang G, Kwok N (2015) Image contrast enhancement for brightness preservation based on dynamic stretching. Int J Image Process (IJIP) 9(4):241
16. Roopa H, Asha T (2016). Segmentation of X-ray image using city block distance measure. In: 2016 International conference on control, instrumentation, communication and computational technologies (ICCICCT). IEEE, pp 186–189
17. Ryzhkov MD (2015) Knee cartilage segmentation algorithms: a critical literature review. Master's thesis
18. Scott D, Kowalczyk A (2007) Osteoarthritis of the knee. BMJ Clin Evid 2007:1121
19. Shamir L, Ling SM, Scott Jr, WW, Bos A, Orlov N, Macura TJ,… Goldberg IG (2009) Knee X-ray image analysis method for automated detection of Osteoarthritis. IEEE Trans Biomed Eng 56(2):407–415
20. Wahyuningrum RT, Anifah L, Purnama IKE, Purnomo MH (2016) A novel hybrid of S2DPCA and SVM for knee osteoarthritis classification. In: 2016 IEEE international conference on computational intelligence and virtual environments for measurement systems and applications (CIVEMSA). IEEE, pp 1–5

Plant Disease Detection Based on Region-Based Segmentation and KNN Classifier

Jaskaran Singh and Harpreet Kaur

Abstract The plant disease detection is the technique which can detect disease from the plant leaves. The plant disease detection has various steps which are textural feature analysis, segmentation, and classification. This research paper is based on the plant disease detection using the KNN classifier with GLCM algorithm. In the proposed method, the image is taken as input which is preprocessed, GLCM algorithm is applied for the textural feature analysis, k-means clustering is applied for the region-based segmentation, and KNN classifier is applied for the disease prediction. The proposed technique is implemented in MATLAB and simulation results show up to 97% accuracy.

1 Introduction

Agricultural area is the only area through which the food requirements of complete human race are being served. In India, around 70% of the total population relies on agriculture and grow several kinds of fruits and vegetable crops within their fields. However, there is a need for high technicality while cultivating crops that are of optimum yield and quality. It is thus important to diagnose the diseases within plants [1]. The identification of diseases is a difficult process due to which the farmers face lots of issues. A research study in which the contagious diseases of the plant are studied is known as potato plant pathology. In order to upgrade the agricultural areas, the disease detection application is applied. Accurate treatment advices are provided through disease management. A computer, digital camera as well as application software are the three components of machine vision system today. Within the application software, numerous algorithms are integrated. In biology, there are several applications that include digital image processing and image analysis technology

J. Singh
ECE Department, Chandigarh University, Gharuan, India

H. Kaur (✉)
CSE Department, Chandigarh University, Gharuan, India
e-mail: harpreet8307@gmail.com

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_154

in order to propose advances within the microelectronics as well as computer systems [2]. The various issues that are relevant to traditional photography are also resolved here. The images that are collected from the microscopic to telescopic ranges are enhanced and analyzed with the help of this tool. Within biology, there are several applications involved. For less production of potatoes, there are numerous factors that are responsible such as early and late blights, insect damage, and roll viral diseases. Within image analysis and pattern recognition, image processing is the initial step and the most important one as well [3]. The final results of the analysis are determined through this step. The process through which an image can be partitioned into disjoint regions is known as image segmentation process.

2 Literature Review

Prakash et al. (2017) studied several image processing techniques in order to identify diseases in plants [4]. In order to identify the leaf diseases and classify them, the image analysis and classification techniques are to be implemented as an aim in this paper. There are four different parts of the proposed framework. They include image processing, segmentation of leaf using k-means clustering in order to identify the diseased areas, feature extraction, and classification of diseases. Using statistical Gray-Level Co-Occurrence Matrix (GLCM) features and classification, texture features are extracted in this approach. The Support Vector Machine (SVM) classifier is used within this proposed mechanism in order to provide better extraction of features.

Kaur et al. (2017) presented a comprehensive study related to several diseases identified within the fruits [5]. In order to minimize the time required to identify the diseases within the fruits, an automated approach is generated for identifying the diseases within fruits. The image is distorted due to the noise. Within this case, there is elaboration of denoising mechanism. As per the analysis of this study, most of the fruit crops are being affected by blight disease. The image of distorted lead is used in order to perform analysis such that the disease can be identified within these fruit crops. In order to identify the diseases within the initial stage, the singular valued analysis is used within the image processing techniques.

Dhaware et al. (2017) proposed a method which can identify the automatic leaf unhealthiness classification from leaves by using image processing [6]. Practical requisition is used here for applying the system because the images are forwarded directly as the farmers make the least efforts. Thus, the farmers that do not make much efforts are provided with advises through this technique. The plant leaves image is seized by the farmer by using the mobile camera. Without making any additional inputs, this image is forwarded to the DSS approach.

Padol et al. (2016) proposed the method that utilized SVM classification technique for identification and classification of diseases present within grape leaves [7]. Within image processing, methods such as resizing, thresholding as well as Gaussian filtering are applied. In order to execute segmentation on the leaf, k-means clustering technique is utilized in which texture and color features are utilized. Further, in order to identify the type of leaf disease, SVM classification technique is utilized. Downy mildew and Powdery mildew are the two different classes of grape leaves that were utilized within this experiment. For both of these categories, an average of 88.89% of accuracy is provided by this proposed system.

Rajan et al. (2016) proposed a mechanism that utilized image processing in order to identify pests within the crops [8]. Upon different images gathered from various sources, tests were performed using this system. At an early stage, it is possible to identify the pests within crops through this proposed method. Thus, the usage of pesticides in agricultural fields can be minimized here through which costs and environment both can be saved. As per the evaluations made, this system is termed as simple and efficient for various applications. In comparison to other manual systems, the time and accuracy level achieved through it is also enhanced.

3 Support Vector Machine

The set of relevant supervised learning mechanisms which are utilized in order to perform classification and regression are known as Support Vector Machines (SVMs). The empirical classification error can be minimized and geometric margin can be maximized simultaneously with the help of SVM. Due to this property, this is also known as Maximum Margin Classifier and it is also based on the Structural Risk Minimization (SRM). Toward the higher dimensional space in which maximal separating hyperplane is generated, the input vector is mapped by SVM. In order to partition the data, two parallel hyperplanes are generated on each side of the hyperplane. The hyperplane that maximizes the distance amongst two parallel hyperplanes is known to be the separating hyperplane [9]. It is assumed here that the larger is the margin or distance amongst these parallel hyperplanes, the better is the generalization error of the classifier. The data points that are considered here are in the form as given in Eq. (1).

$$\{(x1, y1), (x2, y2), (x3, y3), \ldots., (xn, yn)\} \tag{1}$$

Here, $yn = 1/-1$ is a constant that is used to represent the class in which point xn belongs. The numbers of samples involved are represented by n. The p-dimensional real vector is each xn. In order to guard against the variables that have larger variance, scaling is important. With the help of partitioning the hyperplane, the training data is viewed here which includes

$$w.x + b = 0 \tag{2}$$

Here, "b" is a scalar values and the p-dimensional vector is represented by "w". Perpendicular to the separating plane, the vector "w" is indicated. The margin is increased by adding the offset parameters "b". The hyperplane is forced to pass the origin in case "*b*" is not present due to which the solution can be restricted. There is an involvement of parallel hyperplanes here as well which can be described by the equations below

$$w.x + b = 1 \tag{3}$$

$$w.x + b = -1 \tag{4}$$

4 Proposed Methodology

The technique is proposed for the plant disease detection. The plant disease detection techniques consist of the following phases:

1. Preprocessing phase: In the first phase, the image is taken as input and the input image is converted to the gray scale image.
2. Textural feature analysis: The textural features of the input image are analyzed using GLCM algorithm. This algorithm, statistical texture analysis is performed on the observed combinations of intensities for calculating texture features. This calculation is done using the intensities at particular positions that are relevant to each other within an image. There are first-order, second-order and higher order statistics classified on the basis of the number of intensity points or pixels present within each combination. The method through which the second-order statistical texture features can be extracted is known as Gray-Level Co-occurrence Matrix (GLCM) method [10]. In numerous applications, this technique has been applied. Within an image that has number of rows and columns that are equal to the number of gray levels is known as GLCM matrix. The relative frequency through which two pixels are partitioned through a pixel distance (Δx, Δy) and their intensities are i and j for an element is defined as $P(i, j|\Delta x, \Delta y)$. Amongst the gray levels i and j at "d" displacement distance and (θ) angle, the matrix element $P(i, j \mid d, \theta)$ can be considered to have second-order statistical probability values. For an input image that has G gray levels from 0 to G–1 and $M \times N$ neighborhood, the intensity at sample m is considered as $f(m, n)$ for neighborhood line n. Hence,

$$P(i, j|\Delta x, \Delta y) = WQ(i, j|\Delta x, \Delta y) \tag{5}$$

Here, (Table 1)

$$W = \frac{1}{(M - \Delta x)(N - \Delta y)} \tag{6}$$

$$Q(i, j|\Delta x, \Delta y) = \sum_{n=1}^{N-\Delta y} \sum_{m=1}^{M-\Delta x} A \tag{7}$$

and

Table 1 Mathematical formulation of features [12]

Feature	Mathematical formulation
Contrast	$\sum_i \sum_j (i-j)^2 g_{ij}$
Correlation	$\frac{\sum_i \sum_j (ij) g_{ij} - u_x u_y}{\sigma_x \sigma_y}$
Energy	$\sum_i \sum_j g_{ij}^2$
Homogeneity	$\sum_i \sum_j \frac{1}{1+(i-j)^2} g_{ij}$
Mean, M	$\sum_{i=0}^{L-1} g(i) P(g(i))$
Standard Deviation, S	$\sqrt{\sum_{i=0}^{L-1} (g(i)-M)^2 P(g(i))}$
Skewness	$\frac{1}{S^3} \sum_{i=0}^{L-1} (g(i)-M)^3 P(g(i))$
Entropy	$\sum_{i=0}^{L-1} P(g(i)) \log_2 P(g(i))$
Kurtosis	$\frac{1}{S^k} \sum_{i=0}^{L-1} (g(i)-M)^k P(g(i))$
RMS	$\sqrt{\frac{1}{L*L} \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} (g(i,j)-I)^2}$

$$A = \begin{cases} 1\, if\ (m,n) = 1 \text{ and } f(m+\Delta x), n+\Delta y) = j \\ 0 \quad \text{elsewhere} \end{cases} \tag{8}$$

5 Region-Based Segmentation

In this phase, the k-means clustering is applied for the region-based segmentation. The k-means clustering algorithm consists of the following phases:

1. Input data: The image which is taken as input is considered as the data on which segmentation needs to implemented.
2. Calculate arithmetic mean: The arithmetic mean of the input data is calculated which defines centroid point of the data.
3. Formation of segments: The Euclidean distance is calculated for the central point and points which have similar distance is clustered in one cluster and other in the second.
4. Selection of ROI: The segmented image is selected as ROI which is given input to next phase of classification.

6 Classification of Disease

The step of classification will classify the input image into defined disease. The ROI which is selected in the last step is taken as input. The KNN classifier is a technique that is used within the pattern recognition process to classify the objects on the basis of closest training examples present within the feature space is known as K-Nearest Neighbor algorithm (KNN). In this instance-based learning algorithm, there is only local approximation of the function and until classification, all the computation is held upon. Amongst all the machine learning algorithms, KNN is the simplest one. Through the majority vote of neighbors, an object is classified [11]. The class that is most common amongst its k-nearest neighbors is assigned that object. The object is directly assigned to the class of its nearest neighbor in the case when $k = 1$. For both classification and regression predictive issues, NN is utilized. The ease to interpret output; compute the time; and the predictive power available are the three important aspects that are considered while evaluating any technique. By default, Euclidean distance is used by $knn()$ function and the following equation is used to calculate it.

$$D(p, q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 \cdots + (p_n - q_n)^2} \tag{9}$$

Here, the subjects that have "n" properties and are to be compared are represented as "p" and "q".

KNN is also computed by deciding the number of neighbors that can be selected here which is represented by parameter "k". The performance of KNN algorithm is affected on large scale by the choice of value of k. The variance generated due to random error can be minimized by maximizing k. The balance amongst overfitting and underfitting can be maintained on the basis of the selection of appropriate value of k. The output of the KNN classifier will give the disease name.

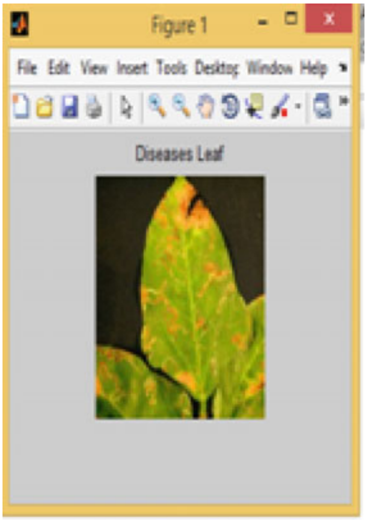
7 Results and Evaluation

From the publically accessible dataset "Plant Village", a database of around 40 images of potato leaves is gathered and used in order to evaluate the proposed algorithm. Image of 10 healthy leaves and 30 diseased leaves are present within the database used for conducting experiments. The complete database is partitioned into 2 different sets during the experiments. There are 24 images (60%) of images present within the training set and 16 images (40%) of images present within the testing set. The KNN that includes linear Kernel is utilized in order to perform classification. Various performance parameters such as accuracy, sensitivity, recall, and F1 score are utilized in order to evaluate the performance of this classification model. There is around 97% of accuracy achieved for classification.

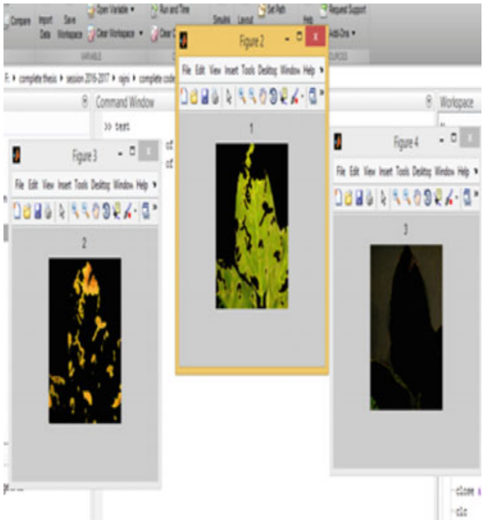
Table 2 shows the performance measures applied. Comparisons are made amongst potato and other species only since there is not much work proposed on disease detection of potatoes. The performance of the KNN classification technique is presented in Table 2.

Table 2 Performance measures of classification

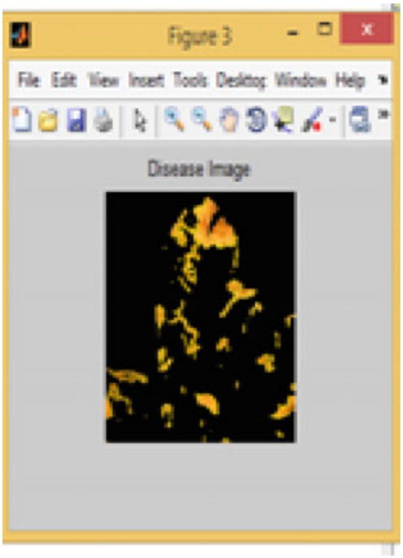
Class	Precision (%)	Recall (%)	F1 score (%)
0: Leaf Minar	89	94	92
1: Mosnic virus	97	93	95
2: White fly	98	98	98
Average/total	97	97	97



(a) Input Image



(b) Segmented Image



(c) Diseased Image

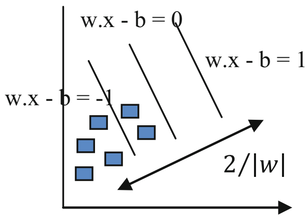


Fig. 1 Maximum margin hyperplanes for an SVM trained with samples from two classes

As shown in Fig. 1, a is the input image for plant disease detection. b Apply k-mean region-based segmentation. c Diseased image.

8 Conclusion

In this research paper, it is concluded that plant disease detection is the approach to detect diseases from the plants. In this work, the GLCM algorithm is applied for the textural feature analysis, k-means clustering is applied for the region-based segmentation, and KNN classifier is applied for the disease prediction. The simulation of the proposed modal is done in MATLAB and results are shown in form of figures and tables. The simulation results illustrated that accuracy is achieved up to 97% of disease prediction.

References

1. Rastogi A, Arora R, Sharma S, Leaf disease detection and grading using computer vision technology and fuzzy logic. In: 2015 2nd international conference on signal processing and integrated networks (SPIN)
2. Khirade SD, Patil AB (2015) Plant disease detection using image processing. In: 2015 international conference on computing communication control and automation
3. Waghmare H, Kokare R, Dandawate Y (2016) Detection and classification of diseases of grape plant using opposite colour local binary pattern feature and machine learning for automated decision support system. In: 2016 3rd international conference on signal processing and integrated networks (SPIN)
4. Prakash RM, Saraswathy GP, Ramalakshmi G (2017) Detection of leaf diseases and classification using digital image processing. In: 2017 international conference on innovations in information, embedded and communication systems (ICIIECS)
5. Kaur K, Marwaha C (2017) Analysis of diseases in fruits using image processing techniques. In: International conference on trends in electronics and informatics ICEI 2017
6. Dhaware CG, Wanjale KH (2017) A modern approach for plant leaf disease classification which depends on leaf image processing. In: 2017 international conference on computer communication and informatics (ICCCI—2017)
7. Padol PB, Yadav AA (2016) SVM classifier based grape leaf disease detection. In: 2016 Conference on advances in signal processing (CASP)

8. Rajan P, Radhakrishnan B, Suresh LP (2016) Detection and classification of pests from crop images using support vector machine. In: 2016 international conference on emerging technological trends (ICETT)
9. Duan X, Zhao T, Li T, Liu J, Zou L, Zhang L (2017) Method for diagnosis of on-load tap changer based on wavelet theory and support vector machine. In: 2017. In: The 6th international conference on renewable power generation (RPG)
10. Parvez A, Phadke AC (2017) Efficient implementation of GLCM based texture feature computation using CUDA platform. In: International conference on trends in electronics and informatics, ICEI
11. Zhang Zhongheng (2016) Introduction to machine learning: *k*-nearest neighbors. Ann Transl Med 4(11):218
12. Islam M, Dinh A, Wahid K (2017) Detection of potato diseases using image segmentation and multiclass support vector machine. In: 2017 IEEE 30th canadian conference on electrical and computer engineering (CCECE)

Analyzing e-CALLISTO Images: Sunspot Number, 10.7 cm Flux and Radio Solar Bursts

R. Sreeneebus, Z. Jannoo, N. Mamode Khan, C. Monstein and M. Heenaye-Mamode Khan

Abstract This paper investigates on the sun's activity based on sunspot images using time series of sunspot numbers and absolute 10.7 cm flux that would also enable to establish relationship between these indices and provide possible forecasts. Moreover, these indices are also compared with the monthly number of solar bursts detected by the Mauritian Radio Telescope which forms part of the e-CALLISTO network.

1 Introduction

This study provides an insight on the sun path with respect to the different long existing indices which are the sunspot images which as the name suggests are the number of spots on the sun and the 10.7 cm radio flux which the waves detected on earth in the wavelength of 10.7 cm coming from the sun and is also referred to as $F_{10.7}$ and compare these two. Once this is done, we will study the solar burst data from a rather recent project known as the Compound Astronomical Low cost Low frequency Instrument for Spectroscopy and Transportable Observatory, e-CALLISTO for short.

R. Sreeneebus (✉) · Z. Jannoo · N. Mamode Khan
Department of Statistics and Economics, University of Mauritius, Moka, Mauritius
e-mail: ravi.sreeneebus1@umail.uom.ac.mu

Z. Jannoo
e-mail: z.jannoo@uom.ac.mu

N. Mamode Khan
e-mail: n.mamodekhan@uom.ac.mu

C. Monstein
Institute for Particle Physics and Astrophysics, ETH Zurich, Zurich, Switzerland
e-mail: monstein@astro.phys.ethz.ch

M. Heenaye-Mamode Khan
Department of Software and Information Systems, University of Mauritius, Moka, Mauritius
e-mail: m.mamodekhan@uom.ac.mu

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_155

The e-CALLISTO aims to build solar radio spectrometers to observe solar bursts within the radio frequency and since 2009, Mauritius has been contributing to it. The aim of this paper is to use the data gathered by the Mauritian radio observatories at Bras d'Eau and compare them with the existing indices that measures the activity of the sun like the sunspot number and 10.7 cm flux in order to explore the relationship between them and perhaps even predict the number of solar bursts that will occur in the future. The e-CALLISTO has been put in place by the Institute for Astronomy, ETH Zurich, and FHNW Windisch, Switzerland spearheaded by Christian Monstein has data as early as 2002.

1.1 Sunspot Number

Wolf's sunspot number is given by the following formula [1].

$$R = k(10\,g + s) \tag{1}$$

where

R Sunspot number,
g Total number of sunspot groups visible,
s Total number of sunspots,
k A scaling factor.

The data for this index is available at the Sunspot Index and Long-term Solar Observations (SILSO) website. This data is essentially in the form of a time series and can be analysed as such. In order to do a forecasting of the data, a simple Auto Regressive-Moving Average model can be applied. In fact various such models have been proposed. The first of which was an ARMA (2, 0), then ARMA (8, 1) and ARMA (9, 0) [2].

While the sunspot number might be the oldest index for solar activity, it is not the only one. Another popular way of monitoring the sun's activity is by looking at the waves emitted by the sun in the radio frequency.

1.2 The 10.7 cm Solar Radio Flux

The 10.7 cm solar radio flux also referred to as $F_{10.7}$ is an indication of solar activity. It is the average amount of radiation emitted in the 2800 MHz frequency or 10.7 cm wavelength and while the name this index is flux, it really is a flux density [3]. The data for the $F_{10.7}$ is available on the NOAA website in the forms of daily and monthly data. These data are taken from The National Research Council Canada Dominion Radio Astrophysical Observatory in Penticton. A number of versions for the $F_{10.7}$ data is available including absolute flux, adjusted flux, and rotational averages, amongst others. For this paper, the absolute flux was used.

Since the data provided by the e-CALLISTO is in the range of radio frequency, this index was chosen to look into the relationship between the two and use this as a basis to then compare them with solar burst data observed by the Mauritian Radio Telescopes (MRT) to find a potential statistical link between them.

The organization of this paper is as follows: In the next section, we display some of the relevant findings in this field illustrating the relationship between the variables followed by a section on the proposed research on the e-Callisto Data. In Sect. 4, the fitting of the different time series models [4–6] is made. The conclusion is presented in the last section.

2 Related Work

Most if not all of them have used the adjusted flux for the relationship however in this study, the absolute flux has been used. The graph showing the difference between the two is shown below.

In Fig. 1, the green part represents the adjflux and the blue part is the absflux. Looking at the graph, the two sets of data are very similar and in fact the correlation between the two is exactly 1 with the sole difference being that the adjflux is higher than the absflux.

When it comes to a relationship between the absolute 10.7 cm flux data and monthly sunspot number was determined. When the $F_{10.7}$ is plotted against the sunspot number, a mostly linear relationship is seen between the two with a slight curve for the smaller values as shown below (Fig. 2).

So using these observation and the past studies [7] as a reference about the monthly adjusted 10.7 cm radio flux data and monthly sunspot number, equations connecting these quantities was developed. When looking at these past studies quite a few ones were suggested by different researchers including linear ones given by equation below,

$$F = 60.1 + 0.932\, N_s \tag{2}$$

where F is the 10.7 cm flux and N_s is the sunspot number.

More complex nonlinear ones which attempted to re-conciliate slight curvature for the initial portions of the graph and the mostly linear features afterwards. Some polynomials [8] are given by

$$F = 65.2 + 0.633\, N_s + 3.76 \times 10^{-3}\, N_s^2 - 1.28 \times 10^{-5}\, N_s^3 \tag{3}$$

and

$$F = 67.0 + 0.572\, N_s + 3.31 \times 10^{-3}\, N_s^2 - 9.13 \times 10^{-6}\, N_s^3 \tag{4}$$

Furthermore equations with exponentials components were suggested [7, 9] given by equations below respectively.

$$F = 0.448\,N_s\left(2 - e^{-0.027\,N_s}\right) + 66 \tag{5}$$

and

$$F = 67 + 0.97\,N_s + 17.6\left(e^{-0.035\,N_s} - 1\right) \tag{6}$$

By looking at all these equations, it formed a basis which was then adapted to the data at hand and 2 equations were developed. A linear one and a non-linear one which used exponential given by equation

$$F = 0.535\,N_s + 60.708 \tag{7}$$

and

$$F = 62 + 0.65\,N_s + 16\left(e^{-0.035\,N_s} - 1\right) \tag{8}$$

and in order to test these two, the values were compared to the different sun cycles recorded starting with cycle 18 to cycle 24 as shown in table (Table 1).

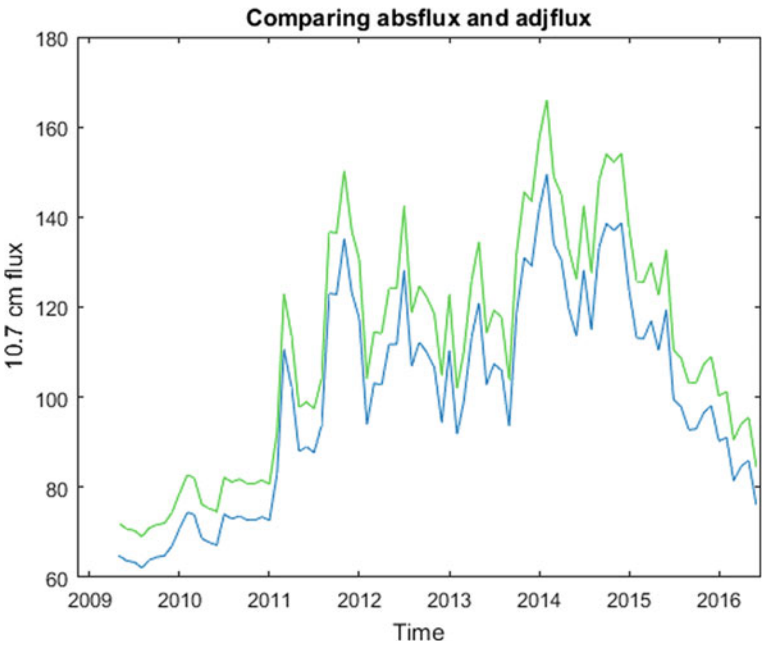


Fig. 1 Absolute flux and adjusted flux

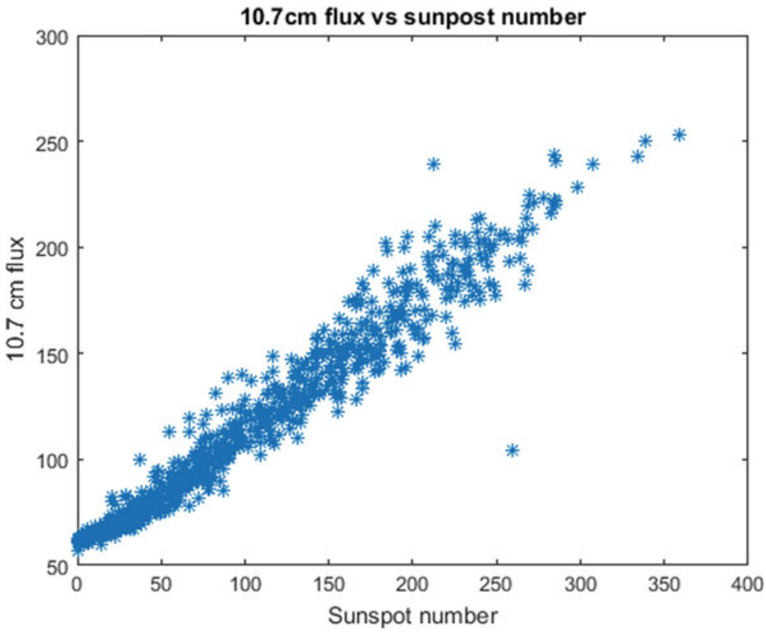


Fig. 2 $F_{10.7}$ versus sunspot number

Table 1 Errors of the 2 methods with respect to sun cycles

Cycle	Linear method		Non linear method	
	RMSE	R^2 (%)	RMSE	R^2 (%)
18	13.23	7.55	13.53	7.90
19	8.32	0.99	8.30	0.99
20	5.77	0.74	7.03	1.09
21	12.90	3.15	14.25	3.85
22	12.63	3.45	8.31	1.50
23	15.02	2.95	13.72	2.46
24	9.46	12.17	9.57	12.47

It can be seen that the error generated for both were close and sometimes the linear model was better while in others, the non-linear model had lower errors. Moreover, by adapting the equation proposed by Tapping [9] about the yearly standard deviation,

$$\sigma_F = 0.23\,\sigma_{N_s}^{1.33} \tag{9}$$

the following equation was developed

$$\sigma_F = 0.23\,\sigma_{N_s}^{1.27} \tag{10}$$

The errors associated with this equation is as follows:
$R^2 = 19.63\%$,
RMSE = 3.90.

3 Proposed Research: e-CALLISTO Data

The e-CALLISTO data is a relatively new a way to look at the sun and allows smaller countries like Mauritius with limited resources to contribute to the study of solar bursts at ground level in the radio frequency.

3.1 Data Gathering

The study of the e-CALLISTO data was first suggested by Dr. Beeharry of the Physics department, at the University of Mauritius. This study provided useful information on the set up and decisions taken by the university to get involved in the project. Furthermore, the data was collected by the MRT telescopes in the year 2009, 2010 and the first 3 months of the year 2011. The data provided was provided in the form shown in Table 2.

For the rest of the data, excel files containing time stamps of the different important activities detected were provided and to analyse these data, the database of the project's website corresponding to the MRT files and at those particular times were consulted.

These data known as spectrogram were available in two forms, namely FITS and image files in the PNG format. The most common way to analyse this type of data is to use the FITS files and with the help of a FITS viewer but due to some problems, another method was devised. A free app on google chrome known as Desmos graphing calculator https://www.desmos.com/calculator was used whereby each PNG file was uploaded and the start time and stop time of each burst was determined as well as the burst frequency range, shown in Table 2 as high and low.

The user interface of this program is seen in Fig. 5 with one of the e-CALLISTO file uploaded to it.

This method being a bit out of the ordinary presents itself with some limitations in the fact that it is not as accurate as the FITS viewers which allow to pin point

Table 2 Example of MRT data

Date code	Burst duration time		Burst description			Burst frequency range	
	Start	Stop	Type	Subtype	Intensity	High	Low
05/01/2009	6:18:50AM	6:19:09AM	III	G	2	116	45

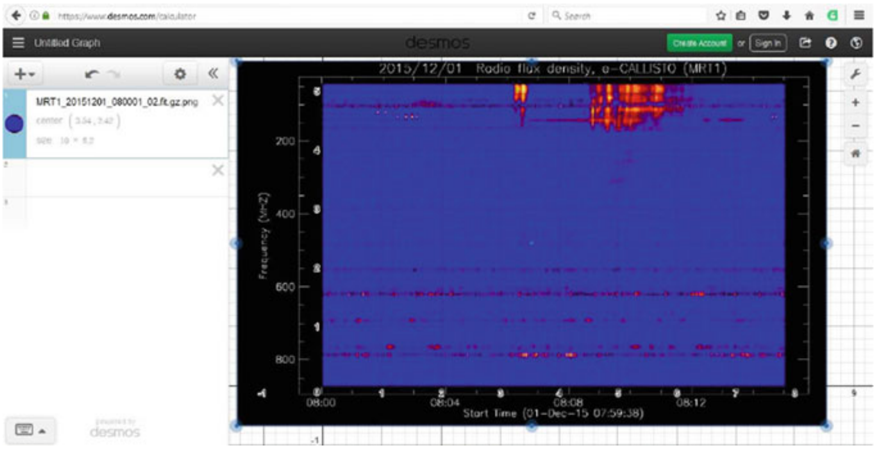


Fig. 3 User interface of Desmos graph calculator

Table 3 Compiled burst data

Date code	Burst duration time			Burst description	Burst frequency range		
	Start	Stop	Duration	Type	High	Low	Burst length
05/01/2009	6:18:50AM	6:19:09AM	16 s	III	116	45	71

exact locations on picture returning their exact coordinates and errors exist specially in identifying the length of the different bursts which is represented in Fig. 3 by the header of high and low and as the name would suggest are the maximum and minimum points respectively of the solar bursts. The error estimated for each of these entries would be of ±5 MHz and that for the time one will about ±3 s.

With the help of these entries and the burst catalog published by Monstein [10], the different bursts were identified and the Desmos calculator and the data after everything is compiled together is shown in a Table 3.

The duration was calculated by subtracting start from stop and burst length was from subtracting low from high.

3.2 Identification of the Different Types of Radio Solar Bursts

Type I

The type I bursts known as noise storm bursts are generally occur below 350 MHz with a length can be very small, with only a few MHz but others of around 30 MHz have also been detected. Their duration most often occurs in the range of 1 s to 1 min

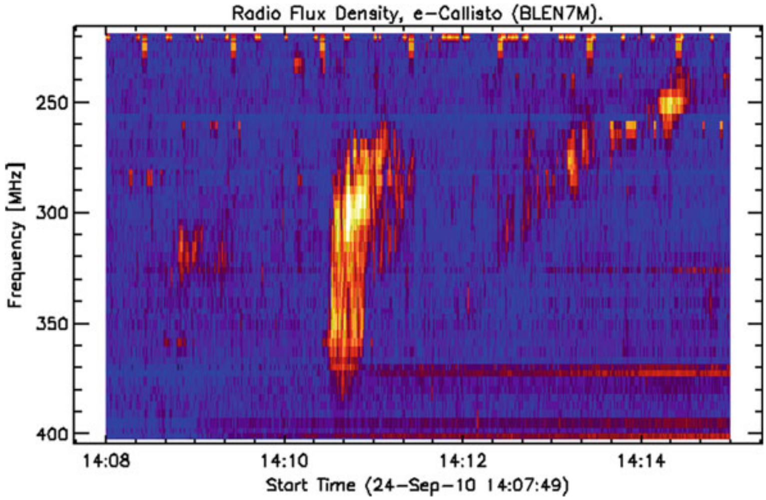


Fig. 4 Type I solar burst or noise storm

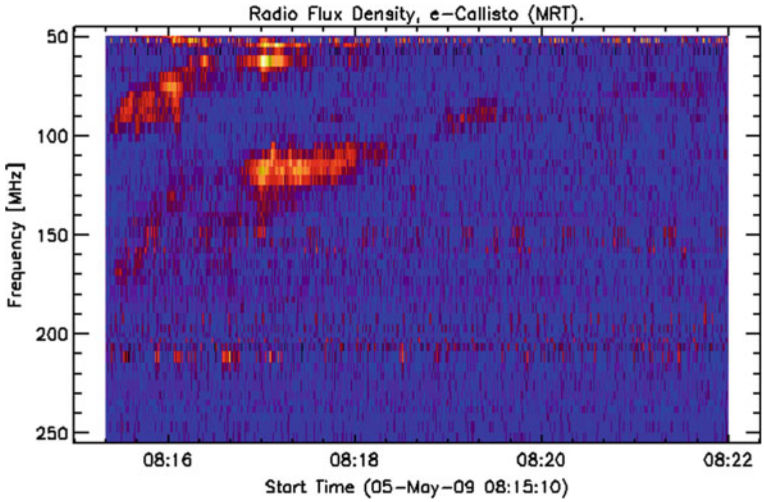


Fig. 5 Type II solar burst with fundamental and harmonic signature

[10]. An example of a type I burst can be seen in Fig. 4. The burst is as a yellow and red spot and the biggest one is the burst in question.

Type II

Slow drift bursts on the images, generally appear long and narrow [10]. Since they generally occur in massive bursts known as CMEs, their duration and length can be quite varied. In Fig. 5, the burst can be seen at the top left.

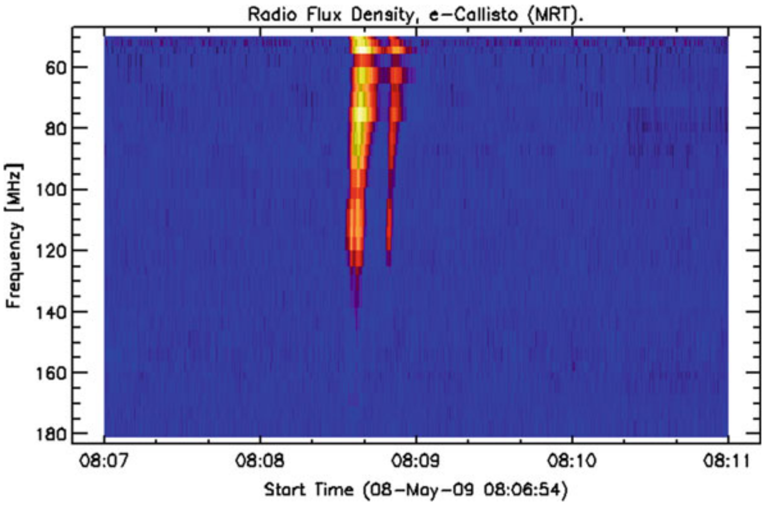


Fig. 6 Type III solar burst

Type III
Fast drift bursts are the most common bursts, and they generally occur at the top of the spectrogram [10]. They may occur individually or in groups. Most often, they last for a few seconds but large groups can go in the minute range. In Fig. 6 at the top we can see a couple of type III bursts.

Type IV
These are the biggest bursts and their duration from hours to even days [10]. Figure 7 shows a type IV burst.

Type V
The type V bursts are closely related to the type III bursts and appear as a flag [10]. Figure 8 shows a type V burst.

If we compare Figs. 6 and 8, you can notice that they are very similar, with the type V having a slightly longer duration.

When all the observations are accounted for, the total number of bursts detected by the Mauritian radio telescopes between May 2009 and June 2016 are 2390 in number and these were divided in terms of type as follows:

Type I: 154.
Type II: 104.
Type III: 2076.
Type IV: 12.
Type V: 37.
Others: 6.

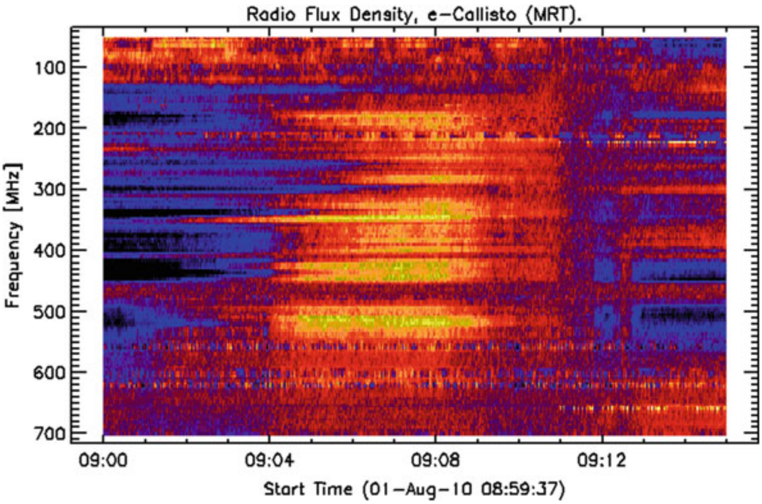


Fig. 7 Type IV solar burst

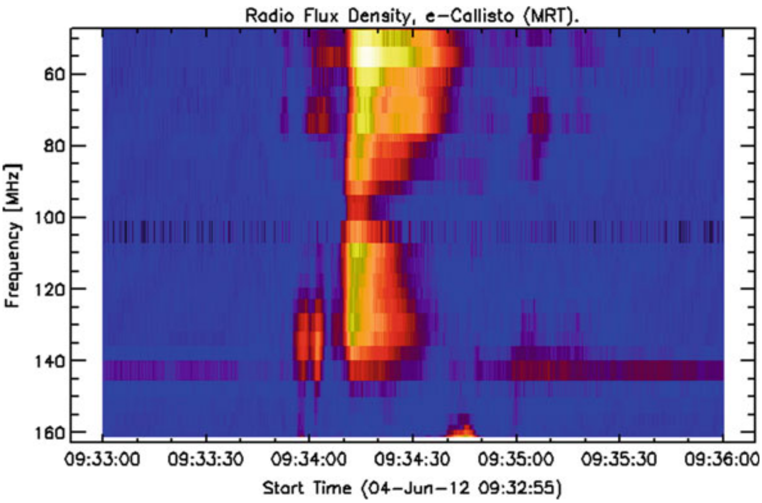


Fig. 8 Type V bursts

4 Analysis and Results: Fitting the Auto Regressive—Moving Average Modelling

4.1 Monthly Sunspot Number

4.1.1 ARMA (9, 0)

$$\begin{aligned} X_t &= 0.5753\,X_{t-1} + 0.1129\,X_{t-2} + 0.0908\,X_{t-3} + 0.0959\,X_{t-4} \\ &\quad + 0.0288\,X_{t-5} + 0.0538\,X_{t-6} - 0.0272\,X_{t-7} - 0.0025\,X_{t-8} \\ &\quad + 0.0362\,X_{t-9} + W_t \end{aligned} \tag{11}$$

The error measures associated with this model are as follows;

AIC = 7805.61,
BIC = 7857.67,
mean error = −0.044,
root mean square error = 25.18,
mean absolute error = 18.10,
mean percentage error = Inf,
mean absolute percentage error = Inf,
mean absolute scale error = 0.93.

4.1.2 ARMA (8, 1)

$$\begin{aligned} X_t &= -0.3438\,X_{t-1} + 0.6805\,X_{t-2} + 0.2068\,X_{t-3} \\ &\quad + 0.2394\,X_{t-4} + 0.1241\,X_{t-5} + 0.0322\,X_{t-6} \\ &\quad - 0.0027\,X_{t-7} + 0.0038\,X_{t-8} + W_t + 0.9690\,W_{t-1} \end{aligned} \tag{12}$$

The error measures associated with this model are as follows;

AIC = 7804.57,
BIC = 7856.63,
mean error = −0.31,
root mean square error = 24.97,
mean absolute error = 18.20,
mean percentage error = Inf,
mean absolute percentage error = Inf,
mean absolute scale error = 0.93 (Table 4).

Table 4 Comparing predicted values with actual values of sunspot number

Month/year	Forecast value for ARMA (9, 0)	Forecast value for ARMA (8, 1)	Actual value
01/2017	27.23	27.39	26.1
02/2017	30.80	28.27	26.4
03/2017	30.17	29.39	17.7
04/2017	30.31	28.90	32.3
05/2017	31.72	31.47	18.9
06/2017	33.08	31.70	19.2
07/2017	34.25	33.90	17.8
08/2017	34.73	33.84	32.6
09/2017	35.14	36.03	43.7
10/2017	36.04	36.05	13.2

4.1.3 Monthly 10.7 cm Radio Flux

ARMA (9, 0)

$$\begin{aligned} X_t = {} & 0.7275\, X_{t-1} + 0.0170\, X_{t-2} + 0.1145\, X_{t-3} + 0.0732\, X_{t-4} \\ & + 0.0018\, X_{t-5} + 0.0546\, X_{t-6} - 0.0035\, X_{t-7} \\ & - 0.0110\, X_{t-8} - 0.0032\, X_{t-9} + W_t \end{aligned} \tag{13}$$

The error measures associated with this model are as follows;

AIC = 6842.87,
BIC = 6894.92,
mean error = −0.25,
root mean square error = 14.08,
mean absolute error = 9.75,
mean percentage error = −1.43,
mean absolute percentage error = 8.01,
mean absolute scale error = 0.99.

4.1.4 ARMA (8, 1)

$$\begin{aligned} X_t = {} & -0.2417\, X_{t-1} + 0.7217\, X_{t-2} + 0.1298\, X_{t-3} \\ & + 0.1815\, X_{t-4} + 0.0684\, X_{t-5} + 0.0538\, X_{t-6} \\ & + 0.0259\, X_{t-7} + 0.0053\, X_{t-8} + W_t + 0.9761\, W_{t-1} \end{aligned} \tag{14}$$

The error measures associated with this model are as follows;

AIC = 6838.28,
BIC = 6890.34,

Table 5 Comparing predicted values with actual values of 10.7 cm flux

Month/year	Forecast value for ARMA (9, 0)	Forecast value for ARMA (8, 1)	Actual value
01/2017	69.99	70.28	67.42
02/2017	72.01	71.90	67.59
03/2017	72.61	72.82	66.56
04/2017	73.18	73.073	73.31
05/2017	73.712	74.01	67.73
06/2017	74.20	74.29	69.47
07/2017	75.01	75.32	72.10
08/2017	75.88	75.69	72.16
09/2017	76.61	76.68	86.90
10/2017	77.28	77.01	68.31

mean error = −0.25,
root mean square error = 14.04,
mean absolute error = 9.74,
mean percentage error = −1.41,
mean absolute percentage error = 7.80,
mean absolute scale error = 0.99 (Table 5).

For both the sunspot number and the 10.7 cm flux. The predictions for the first 2–4 entries seems decent and ones afterward, are to a lesser degree.

4.2 e-CALLISTO Data

4.2.1 ARMA (9, 0)

$$\begin{aligned} X_t = {} & 0.2476\, X_{t-1} + 0.0545\, X_{t-2} + 0.0592\, X_{t-3} + 0.1079\, X_{t-4} \\ & - 0.0497\, X_{t-5} - 0.0892\, X_{t-6} + 0.1344\, X_{t-7} \\ & - 0.0701\, X_{t-8} - 0.2532\, X_{t-9} + W_t \end{aligned} \tag{15}$$

The error measures associated with this model are as follows;

AIC = 799.28,
BIC = 826.27,
mean error = 0.83,
root mean square error = 22.08,
mean absolute error = 17.08,
mean percentage error = Inf,
mean absolute percentage error = Inf,
mean absolute scale error = 0.80.

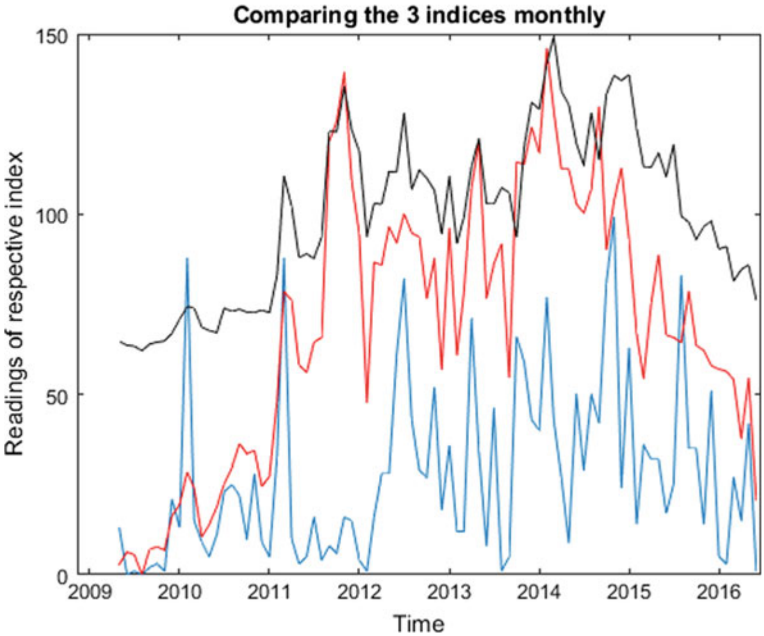


Fig. 9 Monthly time series of all 3 data sets

4.2.2 ARMA (8, 1)

$$\begin{aligned} X_t = & -0.5886\,X_{t-1} + 0.2941\,X_{t-2} + 0.1138\,X_{t-3} \\ & + 0.138\,X_{t-4} + 0.0523\,X_{t-5} - 0.0996\,X_{t-6} \\ & + 0.102\,X_{t-7} + 0.00111\,X_{t-8} + W_t + 0.8964\,W_{t-1} \end{aligned} \quad (16)$$

The error measures associated with this model are as follows;

AIC = 800.87,
BIC = 827.87,
mean error = 0.32,
root mean square error = 22.28,
mean absolute error = 17.39,
mean percentage error = Inf,
mean absolute percentage error = Inf,
mean absolute scale error = 0.81.

Since the data available was not very extensive, all the data was used to fit the models and if a forecast were to be done, there would be no real data to compare with as the process of analysis of the number if bursts detected per month is quite time consuming.

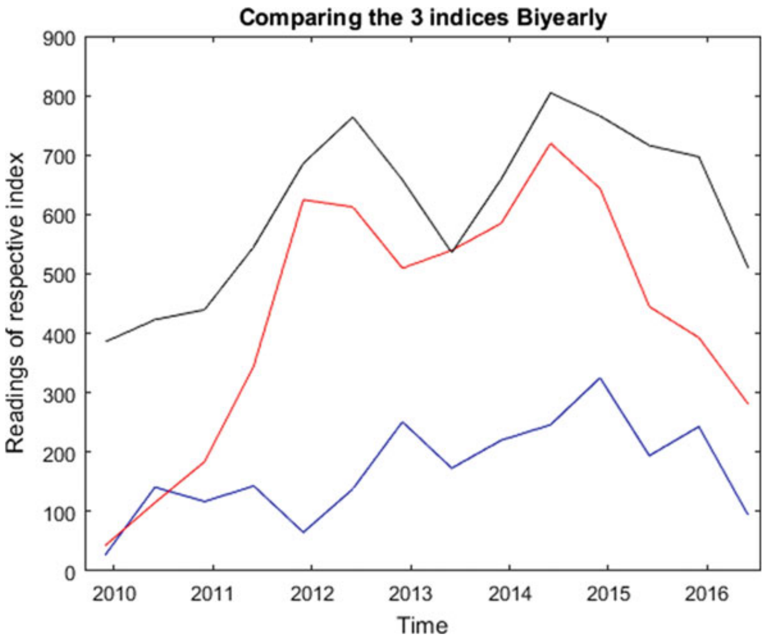


Fig. 10 6 month interval time series of all 3 data sets

4.3 Relationship Between the Solar Burst Data and the 2 Indices

When comparing the 3 data sets for the monthly period between May 2009 till June 2016 is shown below.

In Fig. 9, the red line is the sunspot number, the black line is 10.7 cm flux, and the blue line is the number of burst data detected each month.

Looking at the graph, it is tough for any clear interpretation and the correlation can help us better understand which is as follows:

Correlation between burst data and sunspot number = 0.46.

Correlation between burst data and F10.7 = 0.51.

These numbers are lower than anticipated and one explanation can be because of the data used for burst data was only the data observed in Mauritius. To remedy this, the cumulative data for longer periods that is over periods of 6 months. Therefore the data was divided from periods January to June of a particular year were summed up and that of July and December of that year were summed up and the graph of this particular data sets are shown in Fig. 10.

Similarly to Fig. 9 in Fig. 10 the red line is the sunspot number, the black line is 10.7 cm flux, and the blue line is the number of burst data. The correlation is as follows:

Correlation between burst data and sunspot number = 0.67.

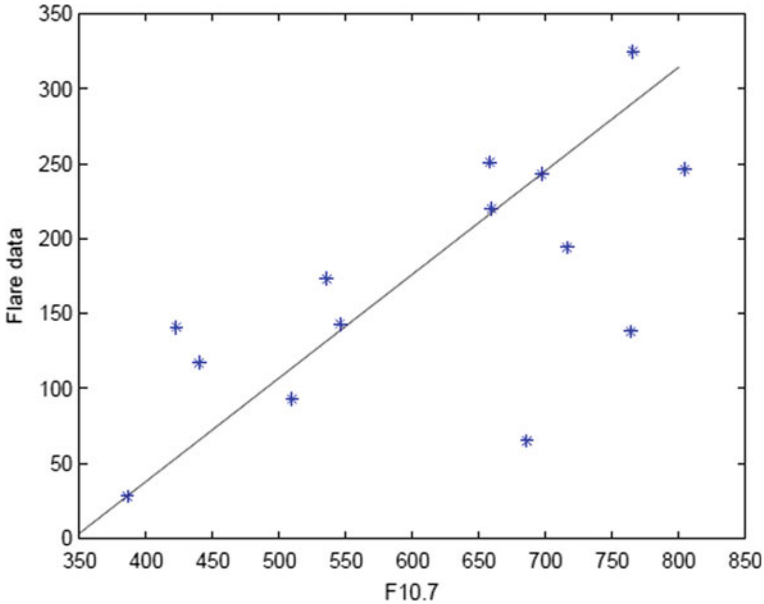


Fig. 11 Relationship between solar burst data and F10.7

Correlation between burst data and F10.7 = 0.74.

Here the correlation is higher and we can observe that it is higher for the burst and $F_{10.7}$. This can be due to the fact that both are in the Radio frequency.

The 10.7 cm flux was used because they are both in the radio frequency and the higher correlation between the two variables.

While a proper equation connecting the two variables could not be derived but a general observation that can be observed is that overall, as the flux increases so does the number of solar bursts as shown in Fig. 11 by the black line.

5 Concluding Remarks and Future Developments

The 11 year cycle was shown using the Fourier transform and time series analysis of the 2 indices that is the sunspot number and 10.7 cm flux as well as that of the solar burst data were carried out with the ARMA (9, 0) being the best model in most cases and the ARMA (8, 1) also being a decent model. The relationship between the absolute 10.7 cm flux and the sunspot number was determined as well as one between the standard deviation connecting the 2 indices. Moreover a general relationship between the sun's activity and that of the solar bursts was shown. As the sun's activity increased so did the number of solar bursts detected.

As for future works, gathering data from the e-CALLISTO using all the stations around the world and make an exhaustive database of all bursts detected to then compare it to the sunspot number and the 10.7 cm radio flux can be a good place to start. Moreover at least the data for a compete sun cycle pertaining to the radio solar bursts will also be very helpful in this endeavour to better understand the behaviour of the sun.

Acknowledgements I would like to thank Dr. Beehary of the Physics department of the University of Mauritius as he was the one who provided the data from the Mauritian Radio Telescope and all those working at MRT for gathering the data.

References

1. Box GE, Jenkins G, Reinsel GC (2008) Time series analysis: forecasting and control. Wiley
2. Werner R (2012) Sunspot number prediction by an autoregressive model. Sun Geosphere 7:75–80
3. Tapping K (2013) The 10.7 cm solar radio flux (F10.7). Space Weather 11:394–406
4. Andersen A, Weiss A (1984) Forecasting: the box-jenkins approach. In: Makridakis et al. (eds) The forecasting accuracy of major time series methods. Wiley. Chichester, England
5. Box GEP, Jenkins G (1970) Time series analysis. Forecasting and Control, Holden-Day, San Francisco
6. Box GEP, Pierce DA (1970) Distribution of the residual autocorrelations in autoregressive-integrated moving-average time series models. J Am Stat Assoc 65:1509–1526
7. Hathway D (2015) The solar cycle. Mail stop: 258–5, NASA ames research center, Moffett field, CA 94035, USA, 1–81
8. Johnson WR (2009) Power law relating 10.7 cm flux to sunspot number. Astrophysics and Space Science 1–8
9. Tapping KF, Valdes JJ (2011) Did the sun change its behaviour during the decline of cycle 23 and 24. Solar Phys 337–350
10. Monstein C (2011) Catalog of dynamic electromagnetic spectra. Phys Astron Electron Work Bench

A Homogenous Prototype Design for IED Detection Using Subsurface Ground-Penetrating RADAR

Alagarsamy Gautami and G. Naveen Balaji

Abstract Security is a prime factor of a nation's peace. Nowadays, terrorists are using Improvised Explosive Devices for their attacks. This paper presents a method to locate, characterize, and identify the IED in landmines using a subsurface technology called Ground-Penetrating RADAR with electromagnetic imaging. By this technique, the homogenous nature of soil, moisture content, dielectric constants of air, soil, and explosives are analyzed. To measure the attenuation and scattering losses of IED underground, a homogenous 2D prototype model is designed with TNT explosives as IED. By simulation, the scattering loss and attenuation are characterized over the step frequency range from 0.5 to 3 GHz. The step frequency shows that the optimal frequency to detect IED in the subsurface scan was 2 GHz. Based on the experimental results, the interference is more in dry soil than wet soil and also the dimension of landmine is directly proportional to the amount of scattering.

1 Introduction

Improvised Explosive Devices is an improvised device incorporating destructive, poisonous, pyrotechnic chemicals which are designed to destroy or to create chaos and catastrophe action. Locating, characterizing, and identifying the IED are a challenging task for any military organization to combat terrorism. IED detection can be performed by bulk detection and trace detection. Bulk detection is an image processing method to detect explosives and its computer vision characteristics, whereas trace detection method implies the chemical and physical characteristics of the explosives materials. Nowadays, wireless technology plays an important role in demining the IED in landmines [1]. IED detection method includes various sensors such as thermal infrared imagery sensors, passive magnetic anomaly sensor,

A. Gautami (✉) · G. Naveen Balaji
SNS College of Technology, Coimbatore, TamilNadu, India
e-mail: aa.gautami@gmail.com

G. Naveen Balaji
e-mail: yoursgnb@gmail.com

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_156

chemical sniffers sensors, ground-penetrating RADAR, and the microwave sensors. To counter this problem of minesweeping and mine clearance, an efficient dual sensor technology called Ground-Penetrating RADAR (GPR) with electromagnetic imaging is addressed in this paper.

2 Background

IED in landmines are buried or implanted on the surface of ground or buried within the ground in a regular pattern or distributed irregularly. It is necessary to understand the construction of IED and its characterization [2]. The common components of IED include switch, initiator, main charge, power source, and a container. Figure 1 shows the construction of IED. There are two major aspects to understand an IED action.

(a) Tactical Characterization.
(b) Technical Characterization.

Tactical characterization is a plot planned to conduct an IED incident, whereas technical characterization includes the hierarchial design of IED with the necessary technical and forensic information of the design.

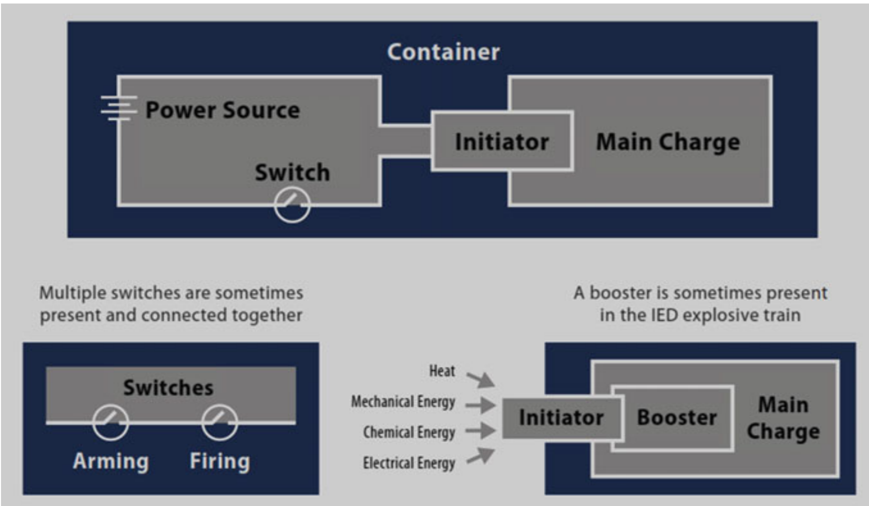


Fig. 1 Construction of IED

Fig. 2 Mılitary vehicle to detect IED using LIBS

2.1 Novel Approaches for IED Detection

The detection of IED depends on various nonhomogenous geometric factors such as depth, size, moisture content, and incident frequency of the soil. By adapting these parameters, some of the novel approaches are listed below.

Spectroscopic approaches: Spectroscopic detection system converts the bomb fumes or vapors to ionized samples in the form of waves [1]. Some of the approaches are:

- Ion Mobility Spectroscopy.
- Mass Spectroscopy.
- Terahertz spectroscopy.
- Infrared spectroscopy.
- Laser-Induced Breakdown spectroscopy (LIBS).
- E-Noses.

Figure 2 shows the military vehicle using LIBS spectroscopic method to detect IED. The spectroscopic methods are accurate and efficient but identification of plastic IED, automated real-time detection, marking, and notification of IED are challenging factors. To overcome these problems, a noninvasive technology called subsurface sensing is modeled to recover IED from landmines [3].

3 Ground-Penetrating RADAR

GPR is an innovative approach for subsurface mapping on collecting RADAR signals from IED. The GPR consists of a transmitter and a receiver which emits electromagnetic waves on IED and detects the target by reflection [4]. Based on the velocity of the reflected signal, soil types, and nature of material, the target is characterized and

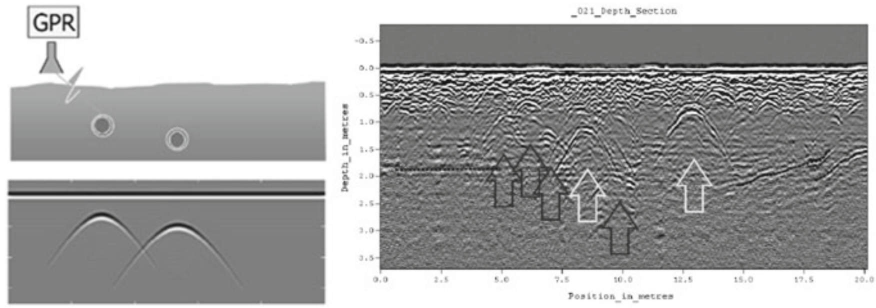


Fig. 3 Target reflection pattern of RADAR from IED

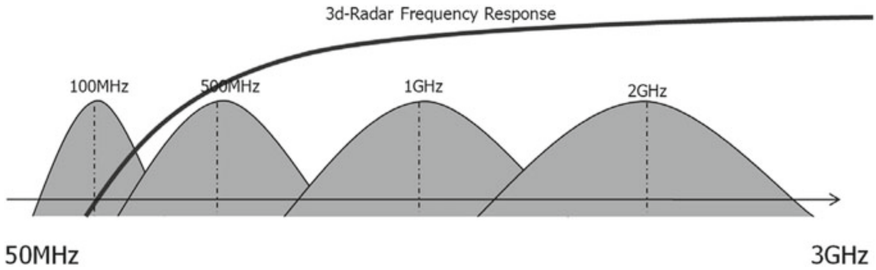


Fig. 4 3D RADAR frequency range coverage for multiple array antenna

identified. Figure 3 indicates target reflection pattern (hyperbola) due to the distance between RADAR and target.

In proper soil conditions, 3D mapping of the geographical area improves good detection than 2D mapping. To perform 3D pattern multiple parallel antennas are mounted in an array on RADAR. Figure 4 shows 3D frequency coverage for multiple array antenna design.

Some of the important features and benefits of 3D GPR are as follows:

- Target detection with real-time analysis for route clearance.
- Array of antennas at variable dimensions can be mounted.
- Detection algorithms for target can be employed.
- Demining system is efficient.

3.1 Data Acquisition Techniques

Step frequency approach is one of the techniques which are used for data collection. In this approach, an incremental frequency steps are transmitted by antenna to the target along the subsurface [5]. The choice of step frequency paves more advantages

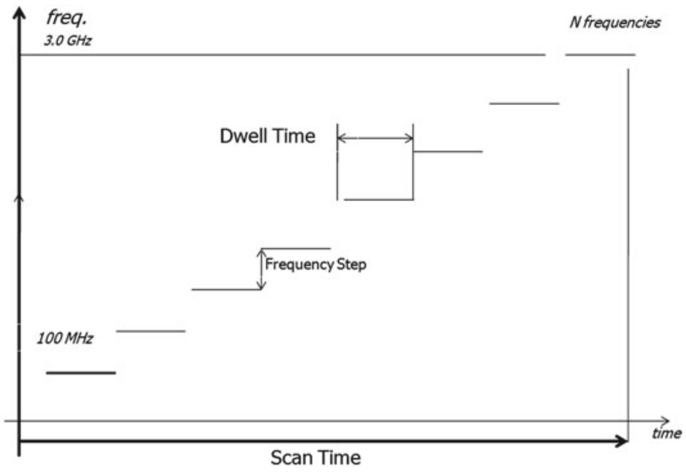


Fig. 5 Graphical representation of step frequency signal

Table 1 The physical parameters of soil and TNT explosives to measure attenuation losses

Material	Relative permittivity	Relative permeability	Conductivity
Air	439.2	1	0
Dry soil	1273 + 31i	2.9	0.004
Wet soil	1756 + 395i 4	4	0.049
TNT	2.9	1	4.8e−4

as shown in Fig. 5, such as maximum integration time, medium resolution is sufficient for greater soil attenuation, and lower interference with radio signals.

3.2 Prototype Design for IED Detection

To determine the GPR experimental results a prototype is designed using COMSOL Multiphysics software with the IED in Landmine scenario. Figure 8 is a 2D prototype model to simulate the parameters and losses to detect GPR actions to locate IED. To promote the modeling the finite element modeling is tested to extract results from the complex environment. Some of the features are adopted in the representation such as two homogenous surfaces which are perfectly matched. These layers absorb all the incident signals from transmitter. The step frequency of 0.5–3 GHz is incident on the surface using EMW module for experimentation. With various geometrical parameters from Table 1 the model is analyzed and simulated to capture scattering loss. Figure 6 shows the 2D prototype model for IED detection.

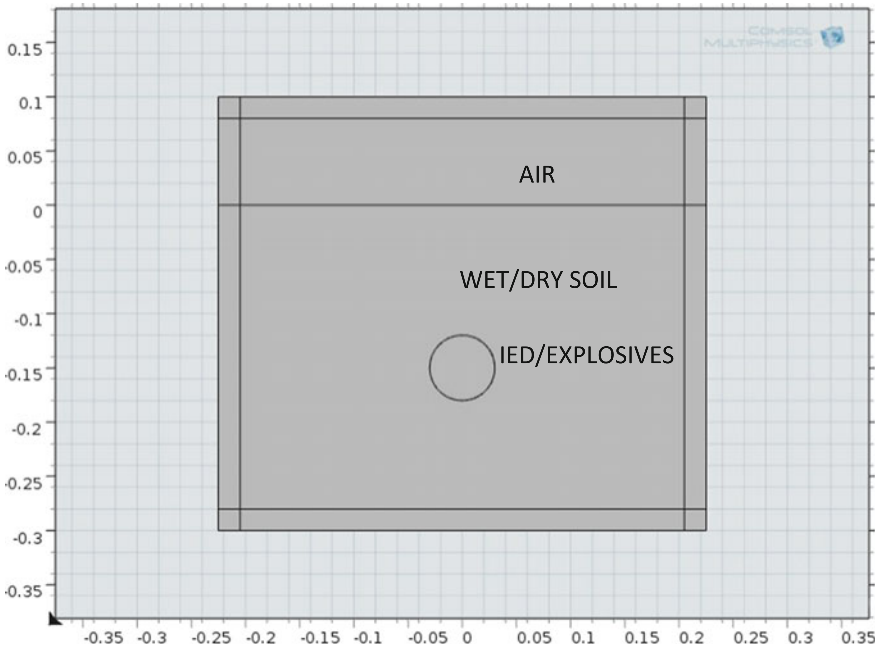


Fig. 6 A 2D model representation for IED detection

3.3 Performance Requirements

The performance factor depends on randomness or probability of detecting IED. Probability Detection (PD) and its false alarm probability (PFA) are the two main factors to measure performance. For example, the maximum performance achieved for detecting IED using handheld metal detector is measured as PD = 1 and PFA = 0. The Receiver Operating Curve (ROC) in Fig. 7 is plotted between probability of false IED to sensitivity of RADAR signal. PD is inversely proportional to PFA [6]. The ROC is used to measure both human and machine performance. True positive (PD) and True negative (PFA) depends on step frequency response. Figure 6 represents the ROC based on TP and TD based on step frequency approach.

The dielectric constant of the soil and explosives is also an important parameter to measure the performance of GPR. For example, the relative dielectric constant of TNT explosives is 2.70 and for RDX it is 2.90. Based on the dielectric constant, the radiated power is measured in the order of milliwatts.

The propagation loss of soil is a performance factor that depends on conductivity of soil and the frequency of the detected signal. The moisture content ensures the relative dielectric constant of soil. From study, the relative dielectric constant for dry to wet soil is 3–16, respectively. Figure 8 represents the spectrum of transmitted and

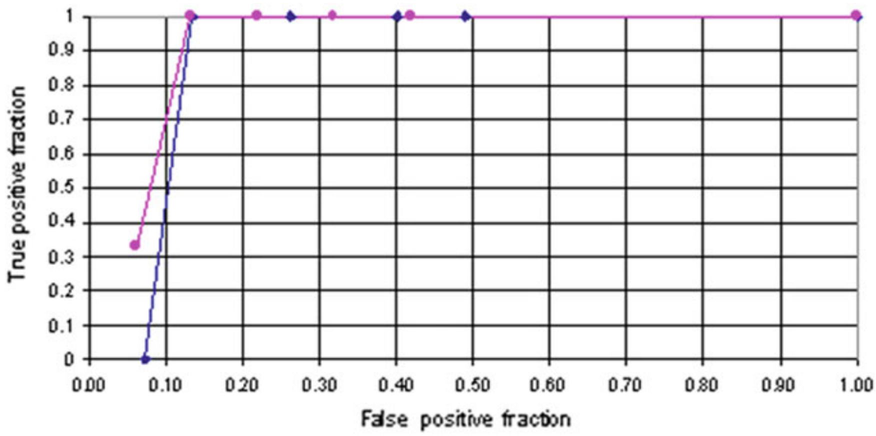


Fig. 7 ROC characteristics (TP vs. TN)

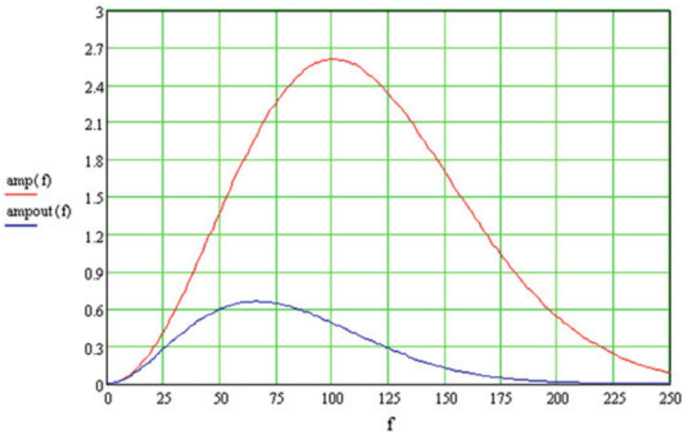


Fig. 8 Transmitted and received signal spectrum after penetrating the lossy soil. (red-transmitted signal, blue-received signal)

received signal after penetrating lossy soil. From Fig. 7, the nature of received signal from lossy dielectric soil is attenuated due to penetration loss.

The attenuation loss is measured using the following formula:

$$\mathrm{La} = 8.686 * 2R * 2\pi\ f\left(((\mu_0 \mu_r \varepsilon_0 \varepsilon_r)/2) * (1 + \tan^{2\delta})^{1/2} - 1\right)^{1/2}$$

where

f	Frequency (Hz)
tan d	Loss tangent of material
ε_r	Relative permittivity of material

ε_0 Permittivity of free space
μ_r Relative magnetic susceptibility of material
μ_0 Magnetic susceptibility of free space
R Range (meters).

From the equation, the performance is featured as follows: the propagation losses are proportional to frequency rates and the near-field regions [7]. The attenuation losses increase with frequency, and it falls in the range 300 MHz–1.5 GHz at receiving end. The attenuation losses will reduce the directivity and gain of multiple array antennas connected to the RADAR.

4 Experimental Results

From the experimental research through simulation, the scattering effects are directly proportional to the IED depth in perfectly matched layer. The step frequency of 0.5–3 GHz is incident on surface using EMW module. From Table 1, the TNT explosive was taken for reference. With the dielectric constant of 2.70, the scattering effects are simulated with the prototype model. From simulation, the dry soil encounters more interference than the wet soil as shown in Fig. 9. Therefore, conductivity is more in the wet soil which can retain waves more than dry soil. The depth of the landmine is proportional to the amplitude of scattered waves. The step of frequencies shows that the optimal frequency to detect IDE in the subsurface scan was 2 GHz.

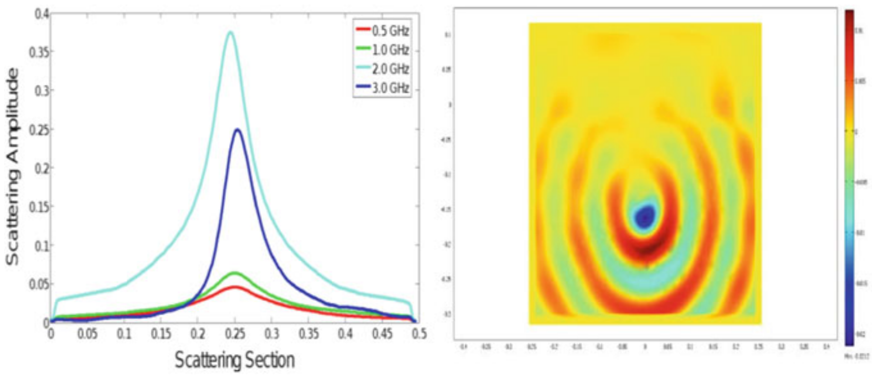


Fig. 9 Scattering amplitude cross at different frequencies for the air/wet soil/dry soil model

4.1 Other Applications of GPR

GPR systems are widely used in demining and detecting IED. It is also used in various capacities with high-resolution electromagnetic imagery such as

- Inspection of Bridge deck.
- Road surveying.
- Delimitation identification.
- Ballast inspection.
- Utility mapping and unmanned missions.
- Archeology.

5 Conclusion

In this paper, an experimental homogenous prototype model for the subsurface detection of IEDs using GPR technology was designed. By developing this prototype model, the nature of radio waves at various dielectric constants is identified. The relation between various parameters which affects the performance of incident radio waves is analyzed. The scattering and attenuation losses of incident waves with step frequency input signal is simulated. This technique provides the limelight to characterize, locate, and identify the IED and explosives all across the globe to counter terrorism.

References

1. Ghazali KH, Jadin MS (2014) Detection improvised explosive devices (IED) emplacement using infrared image: ieee computer society—16th international conference on computer modeling and simulation https://doi.org/10.1109/UKSim.2014.111
2. Muthukumar K, DeviMala E (2011) Detection of IED using nanotechnology: Int J Sci Eng Res 2(12) ISSN 2229-5518
3. Liu R, Wang H: Detection and localization of improvised explosive devices based on 3-axis magnetic sensor array system: 2010 symposium on security detection and information processing: Elsevier. Proc Eng 1877–7058 doi:10.1016
4. Lopera O, Milisavljevie N, Daniels D, Macq B (2007) Time-frequency domain signature analysis of GPR data for landmine identification, advanced ground penetrating RADAR. In: 2007 4th international workshop 27–29 2007, pp 159–162
5. Daniels DJ (2004) Ground penetrating RADAR: ISBN 0863413609: IEEE (RADAR, Sonar and Navigation):year-2004
6. Tripe Wires: Introduction for explosives. Homeland Security Department, USA
7. The White House Countering improvised explosives devices. Washington DC, USA

A Study on Various Deep Learning Algorithms to Diagnose Alzheimer's Disease

M. Deepika Nair, M. S. Sinta and M. Vidya

Abstract Alzheimer's disease (AD) is one of the most frequent types of dementia, which is deterioration in mental ability severe enough to interfere with daily life and gradually affect the human's brain, its capability to learn, think and communicate. The symptoms of AD develop over time and become a major brain disease over the course of several years. To bring out patterns from the brain neuroimaging data, different statistical and machine learning approaches have been used to find the Alzheimer's disease present in older adults at clinical as well as research applications; however, differentiating the phases of the Alzheimer's and healthy brain data has been difficult due to the similarity in brain atrophy patterns and image intensities. Recently, number of deep learning methods has been expeditiously developing into numerous areas, which consist of medical image analysis. This survey gives out the idea of deep learning-based methods which used to differentiate between Alzheimer's Magnetic Resonance Imaging (MRI) and functional MRI from the normal healthy control data.

1 Introduction

The research in Deep learning has been growing fast in the medical imaging field, including Computer-Aided Diagnosis (CAD), medical image analysis, and radiomics. CAD is an expeditiously developing area of research in the medical industry. The recent researchers in machine learning guarantee the enhanced efficiency in detection of disease. Here, the computers are enabled to think by

M. Deepika Nair (✉) · M. S. Sinta · M. Vidya
Department of Computer Science Engineering, Vidya Academy of Science and Technology, Thrissur, Kerala 680501, India
e-mail: deepikamnair1995@gmail.com

M. S. Sinta
e-mail: sintamsuresh@gmail.com

M. Vidya
e-mail: vidya.m@vidyaacademy.ac.in

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_157

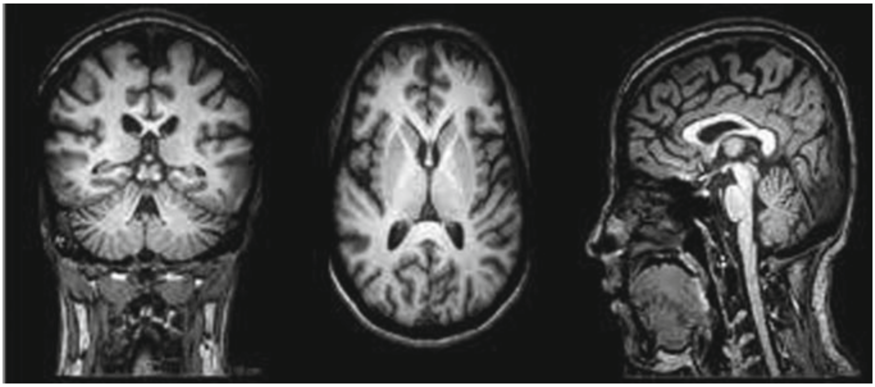


Fig. 1 T1-weighted MRI scans acquired in coronal (left), axial (center), and sagittal (right) planes with 3 T

developing intelligence by learning [1]. There are distinct types of deep learning techniques and which are used to analyze the data sets (Fig. 1).

One of the symptoms of Alzheimer's disease is the enlargement of amyloid plaques among brain nerve cells [2]. To diagnosis the Alzheimer's disease, it is required to know the internal structures hidden in the brain which can be obtained through various types of scanning techniques existed.

These scanning methods in the medical field include Angiography and Computed Tomography (CT), MRI Angiography (MRA), Dynamic CT or Dynamic MRI, Magneto Encephalography (ME), MRI, Flow-sensitive MRI (FSMRI), functional MRI (fMRI), etc. Structural MRI is a chief imaging biomarker in AD as the cerebral atrophy is shown to closely link with cognitive symptoms.

Deep Learning (DL) is a member of machine learning research inspired by the function and structure of human brain, and which aim at discovering multiple levels of distributed representations. Recently, many deep learning procedures have been used to fix conventional artificial intelligence problems. Mainly, there are two groups of deep learning models which are dissimilar with respect to how the data flow over the network. The information in the feed-forward networks flows over the network in just one direction, from the input layer to the output layer. Compared to feed-forward network, recurrent networks have feedback connections that allow the information from past inputs to affect the current output. In the framework of supervised classification problems, the implementation of Deep Learning requires two main steps. The first stage is called training and this phase uses the training set which is a portion of the dataset available to correct the networks parameters to perform the classification. The next step is testing phase, which is used the rest of the subset called as the test set to determine whether model that is trained can correctly predict the new observations class. When the number of available data is less, it is also possible to run the training and testing phases several times on different training and test splits of the original data and then estimate the average performance of the model. This approach is known as cross-validation. The training phase and testing phase is not a unique feature of Deep Learning but are used in conventional ML methods [3].

The paper is organized is as follows. Section 2 describes various deep learning techniques to detect Alzheimer's disease. Section 3 concludes the paper.

2 Related Works

This section reviews existing Deep Learning approaches used to detect Alzheimer's disease from structural and functional MRI scans.

2.1 Autoencoder

An autoencoder is an artificial neural network which can grasp the features in an unsupervised way by minimizing reconstruction errors [4]. The intention of autoencoder is to train a model for the collection of data, with the typical purpose of data compression. The process of training in autoencoder is based on the development of a cost function. The cost function estimates the error through backpropagation and it is rebuilding at the output. An autoencoder constitutes of an encoder followed by decoder. The encoder and decoder can possess multiple layers; nevertheless for simplicity, we consider that all of them have only one layer. The basic structure of autoencoder is shown in Fig. 2.

Debesh Jha and Goo-Rak Kwon Debesh Jha proposed a method based on autoencoder and this framework use structural MRI data provided from Open Access Series of Imaging Studies (OASIS) database [5]. Here use deep learning architecture, which encompass of sparse autoencoders, Scale Conjugate Gradient (SCG), stacked autoencoder, and a softmax output layer to overcome the bottleneck and support the analysis

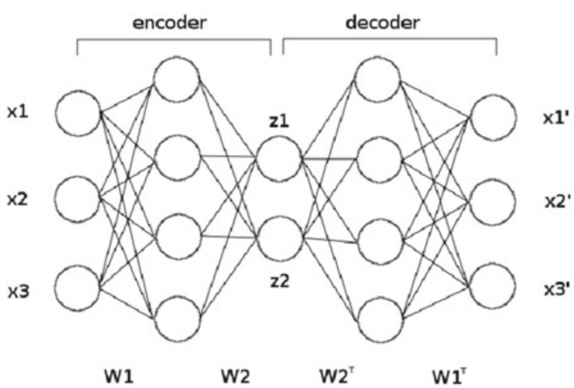


Fig. 2 Basic structure of autoencoder

of AD and normal healthy controls. Compared to the former workflows, this technique requires fewer labeled training examples and minimum prior knowledge and also performs dimensionality reduction and data fusion at the same moment. A performance gain is achieved with the binary classification and gets 91.6% accuracy.

Bhatkoti and Paul [6] studied the effectiveness of the k-sparse Autoencoder (KSA) algorithm in deep learning structure for the diagnosis of Alzheimer's disease. It compared the modified approach to non-modified k-sparse approach in this application. The research used MRI scan data, CSF, and PET images each of 150 patients. MRI images for comparison with research images together with CSF and PET data were obtained from Alzheimer's disease Neuroimaging Initiative (ADNI). The MRI scans were reprocessed by correcting orientation errors and by skull strip to obtain underlying tissues. The images were normalized and smoothened. Patch extraction was then done and masks for different brain subregions were obtained and transformed during registration in the Automatic Anatomical Labeling (AAL) template. A feed-forward convolutional pair predictor neural network was developed. Flattening, concatenation, and sorting were done on feature vectors which were in turn input into feed-forward multi-layer perceptron. Prediction of output was carried out using probability function. Three-dimensional convolutional neural network with 384 input neurons and 200 hidden neurons were used in multi-layer perceptron. Cross-validation algorithm with 20-fold cross-validation was used in training. A practical approach with actual MRI images from patient screening was used in this research and compared with data from ADNI as well as those employed in previous studies, and this method contributes to efficiency of 63.24% early diagnosis of Alzheimer's disease and confirms that KSA enhances the efficiency as 74.05%.

2.2 Convolutional Neural Network

A Convolutional Neural Network (CNN) is made up of convolutional layers, pooling layers, normalization layers and then the fully connected layers. The construction of a CNN is described to yield the advantage of the 2D format of an input image. CNN use little preprocessing operation rather than other image processing operations. The advantage of CNNs is that it cut down the number of parameters with the same number of hidden units and the training of CNN is straightforward. The convolutional layer takes an MxMxR input image, which is corresponding to the height, width, and the number of channels of an image, respectively.

The convolutional layer plays a crucial role in CNN architecture and is the basic building block in this network. The CONV layers parameters contains a group of learnable filters. Spatially, the size of each filter is small but enhance through the complete input volume depth. For each forward pass, it is multiplying the original image pixel values with the values in the filter which is used by CONV layer. These multiplications are all summed up and producing a two-dimensional activation map of that filter. Next, all filters activation maps are stacked and produce output volume. A pooling layer is mostly added in-between subsequent Conv layers. Its function is to

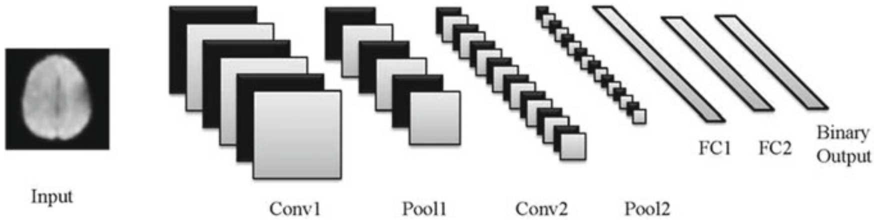


Fig. 3 LeNet-5 architecture

reduce the dimensionality of each feature map but preserve the important information. The pooling Layer operates on each feature map independently and resizes the input spatially. Spatial pooling can be of different types Max, Average, Sum, etc. Recent researchers have developed more successful CNN such as AlexNet, GoogleNet, ResNet, ZF Net, VGGNet, and LeNet. The major problem of constructing ConvNet architectures is the memory restrictions of GPU [7–9].

Sarraf and Tofighi [10] in their paper used CNN deep learning architecture (LeNet) that was trained and tested with huge number of images and classified the AD data from normal control with 96.86% accuracy. The architecture of LeNet-5 is shown in Fig. 3.

Ciprian et al. [11] use DemNet architecture which is a modified version of the 16-layer CNN made by the Oxford University Visual Geometry Group (VGG) for the ImageNet Large-Scale Visual Recognition Challenge (ILSVRC). It is a collection of 13 convolutional layers followed by 3 fully connected layers. This architecture successfully classifies AD and MCI from HC on the Alzheimer's disease Neuroimaging Initiative (ADNI) dataset with an accuracy of 91.85%.

Glozman and Liba [12] propose a method for AD classification using AlexNet architecture includes five convolutional layers and three fully connected layers. Here, the network pretrained on natural images can be fine-tuned to classify neuroimaging data in which the difference between the different classes are very subtle, even for the human eye. This method results suggest that with the available data, the network can learn to classify the two extreme classes (NC vs. AD), but when faced with a three-way classification task, it will not achieve good accuracy.

3 Conclusion

In this paper, a review on different deep learning methods for diagnosis of Alzheimer's disease is discussed. Generally, the disease developing through three stages: Normal control, MCI, and AD. The Alzheimer's disease causes some changes in the brain, and these changes mainly developed on both the structures which are larger and small cells in the brain. Alzheimer's disease affects some parts of limbic system mainly the hippocampus, then the cerebral cortex, finally the brain stem. Most of the methods use MRI images because it is considered as the favored neuroimaging examination for Alzheimer's disease.

Early diagnosis of AD and MCI based on deep learning methods needs only minimal prior knowledge dependency in the model optimization. The advantage of autoencoder technique is it requires fewer labeled training examples and minimal prior knowledge. Compared to autoencoder technique, CNN deep learning architecture which was trained and tested with a huge number of images classified the AD more accurately.

References

1. Bishop C (2006) Pattern recognition and machine learning. Springer-Verlag, New York
2. American Society of Health-System Pharmacists (2001) Alzheimer's disease education and referral center. Am J Health-Syst Pharm 58(9):826
3. Vieira S, Pinaya WHL, Mechelli A, Serif L (2017) Using deep learning to investigate the neuroimaging correlates of psychiatric and neurological disorders: methods and applications. Neurosci Biobehav Rev 74(Part A)
4. Shin HC, Orton MR, Collins DJ, Doran SJ, Leach MO (2013) Stacked autoencoders for unsupervised feature learning and multiple organ detection in a pilot study using 4d patient data. IEEE Trans Pattern Anal Mach Intell 35:19301943
5. JhaD, Kwon G-R (2017) Alzheimer's disease detection using sparse autoencoder, scale conjugate gradient and softmax output layer with fine tuning. Int J Mach Learn Comput 7(1)
6. Bhatkoti P, Paul M (2016) Early diagnosis of alzheimer's disease: a multi-class deep learning framework with modified k-sparse autoencoder classification. IEEE 2016
7. Arel I, Rose DC, Karnowski TP (2010) Deep machine learning a new frontier in artificial intelligence research [research frontier]. Comput Intell Mag IEEE 5(4):1318
8. Jia SE, Donahue J, Karayev S, Long J, Girshick R, Guadarrama S, Darrell CT (2014) Convolutional architecture for fast feature embedding. In: Proceedings of the ACM international conference on multimedia, pp 675678, ACM
9. LeCun Y, Bottou L, Bengio Y, Haffner P (1998) Gradient based learning applied to document recognition. Proc IEEE 86(11):22782324
10. Sarraf S, Tofighi G (2016) Classification of alzheimer's disease using fMRI Data and deep learning convolutional neural networks
11. Ciprian D, Billones Jr, Louville OJ, Demetria D, Earl D, Hostallero D, Prospero, Naval Jr. C (2016) A convolutional neural network for the detection of alzheimer's disease and mild cognitive impairment. IEEE 2016
12. Glozman T, Liba O (2016) Cues: deep learning for alzheimer's disease classification. CS331B project final report, 2016

Performance Analysis of Image Enhancement Techniques for Mammogram Images

A. R. Mrunalini and J. Premaladha

Abstract Mammography is a technique which uses X-rays to take mammographic images of the breast, but identifying abnormalities from a mammogram is a challenging task. Many Computer-Aided Diagnosis (CAD) systems are developed to aid the classification of mammograms, as they search in digitized mammographic images for any abnormalities like masses, microcalcification which is difficult to identify especially in dense breasts. The first step in designing a CAD system is preprocessing. It is the process of improving the quality of the image. This paper focuses on the techniques involved in preprocessing the mammogram images to improve its quality for early diagnosis. Preprocessing involves filtering the image, applying image enhancement techniques like Histogram Equalization (HE), Adaptive Histogram Equalization (AHE), Contrast-Limited Adaptive Histogram Equalization (CLAHE), Contrast Stretching, and Bit-plane slicing; filtering techniques like mean, median, Gaussian and Wiener filters are also applied to the mammogram images. The performance of these image enhancement techniques are evaluated using quality metrics, namely Mean Square Error (MSE), Peak Signal-to-Noise Ratio (PSNR), and Contrast-to-Noise Ratio.

1 Introduction

1.1 Mammogram

Cancer is caused when abnormal tissues multiply uncontrollably. Breast cancer develops in the breast tissue and can be detected using mammography, which is an effective breast imaging technique to detect cancer at an early stage through low-dose X-rays, even before the symptoms are experienced. Mammography requires higher exposure of radiation for detecting any signs of tumor as the breast contains soft tissues. The

A. R. Mrunalini · J. Premaladha (✉)
School of Computing, SASTRA Deemed-to-be-University, Thanjavur, India
e-mail: premi.ph.d@gmail.com

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_158

mammographic lesions can be diagnosed using a Computer-Aided Diagnosis (CAD) system and can be classified according to its severity.

1.2 Image Enhancement

Mammograms are gray scale images with low contrast and hence, it is difficult to visually distinguish between normal and lesion tissues. Mammogram enhancement is an important step to detect abnormalities and increase the reliability of CAD system. Enhancement of mammograms without blurring fine details and edges is a challenging task as it involves highlighting the suspicious regions without affecting the normal breast tissues. Image enhancement enhances the interested regions to detect the defects in a simple and efficient manner.

Preprocessing significantly enhances the image quality by suppressing uninterested areas. It is the basic step and its accuracy determines the probability of success of segmentation and classification. Filters like mean, median, Gaussian, and Wiener are used to remove salt-and-pepper noises from a mammogram image. The results are compared using parameters such as PSNR, MSE, and SNR to determine the better filter.

1.3 Literature Survey

Mean, median, adaptive median, Gaussian, and Wiener de-noising filters are used to remove salt-and-pepper noises, Gaussian noise, and speckle noise from mammogram images and comparison of the performance is done based on PSNR, MSE, and CNR parameters and the better filter is finalized [1]. Histogram Equalization (HE) is an efficient image enhancement technique. Adaptive Histogram Equalization (AHE) is an efficient contrast enhancement technique that is suitable for medical images and other initially nonvisual images. Contrast-Limited Adaptive Histogram Equalization (CLAHE) prevents local overenhancement using an amplitude limiting method [2]. The objective of image processing field is image de-noising. Preprocessing or image de-noising is necessary for noise suppression and image quality improvement [3]. Histogram equalization is an effective technique to achieve contrast enhancement. Digital mammography deals with gray scale images [4]. Enhancing the mammograms helps the segmentation of breast region for automated analysis of digital mammogram images to be done easier. The Region of Interests (ROI) in the image can be determined by eliminating background noise and improving the image quality [5]. The number of false positives can be reduced and the accuracy of determining the abnormalities can be increased through preprocessing hence making segmentation easier [6]. Image preprocessing is essential for training the CNNs. The performance of CNN can be improved particularly through including brightness and contrast variations which are among the preprocessing techniques [7]. Contrast-

Limited Adaptive Histogram Equalization (CLAHE) and morphology methods are used in particular for mammograms [2].

2 Image Enhancement Techniques

Image enhancement techniques accentuate or sharpen image features such as edges or boundaries to make it helpful for analysis. Image enhancement techniques include:

1. Intensity transformation functions
2. Histogram processing
3. Spatial filtering
    - Linear
    - Nonlinear

2.1 Intensity Transformation Functions

Intensity transformation technique is done to increase the contrast between certain intensity values to analyze particular features in an image. Some of the intensity transformation techniques are:

- Contrast stretching
- Bit-plane slicing.

2.1.1 Contrast Stretching

Contrast stretching is a technique where the intensity range of values in an image is stretched to improve the contrast in an image [8]. It is a measure of the complete range

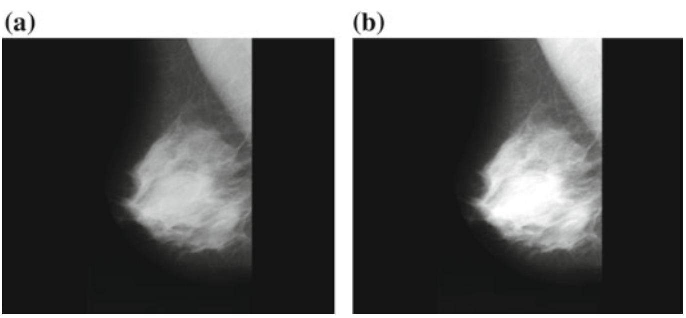


Fig. 1 Contrast stretching http://peipa.essex.ac.uk/pix/mias/all-mias.tar.gz

of intensity values contained within an image and can be calculated by subtracting minimum pixel value from the maximum pixel value. Contrast stretching was applied using MATLAB to the mammogram image obtained from the mini-MIAS database. Figure 1 shows the original image (a) and the enhanced image (b).

2.1.2 Bit-Plane Slicing

Bit-plane slicing is done to convert a gray level image to a binary image. It represents an image with fewer bits and enhances the image by focusing. The contribution made to the total image appearance is highlighted by specific bits. It is based on the assumption that each pixel in the image is represented by 8 bits and the image is composed of 8, 1-bit planes. The least significant bit is composed by plane 1 and the most significant bit by plane 8. The relative importance of each bit can be analyzed through this technique. Bit-plane slicing was applied to the image taken from mini-MIAS database named as "mdb001.pgm". Figure 2a is the original image, (b) is the second bit, (c) is the sixth bit, and (d) is the eighth bit.

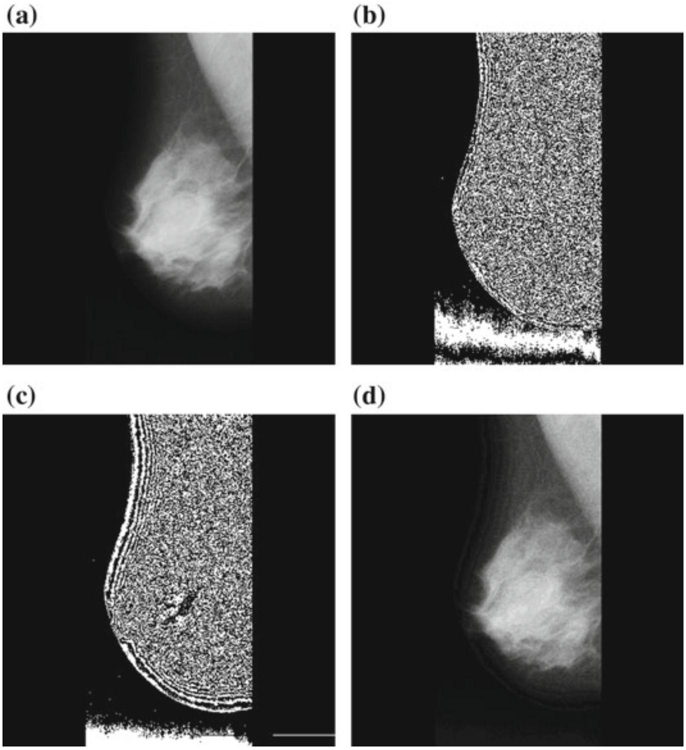


Fig. 2 Bit-plane slicing http://peipa.essex.ac.uk/pix/mias/all-mias.tar.gz

2.2 *Histogram Processing*

Histograms can be used for image enhancement. It is a plot of the frequency of occurrence of an event. It is a spatial domain technique. Image intensities can be adjusted through Histogram Equalization technique to enhance the contrast. There are three Histogram Equalization techniques:

- Histogram Equalization (HE)
- Adjusted Histogram Equalization (AHE)
- Contrast Adjusted Histogram Equalization (CLAHE).

2.2.1 Histogram Equalization (HE)

Histogram Equalization produces an output whose histogram is uniform by altering the input histogram. The various pixel intensities in the output histogram are equally distributed over the entire dynamic range. "histeq" is the command used in MATLAB for Histogram Equalization.

Histogram Equalization was applied to the original image taken from the mini-MIAS database. Figure 3a is the original image, (b) is the histogram of the original

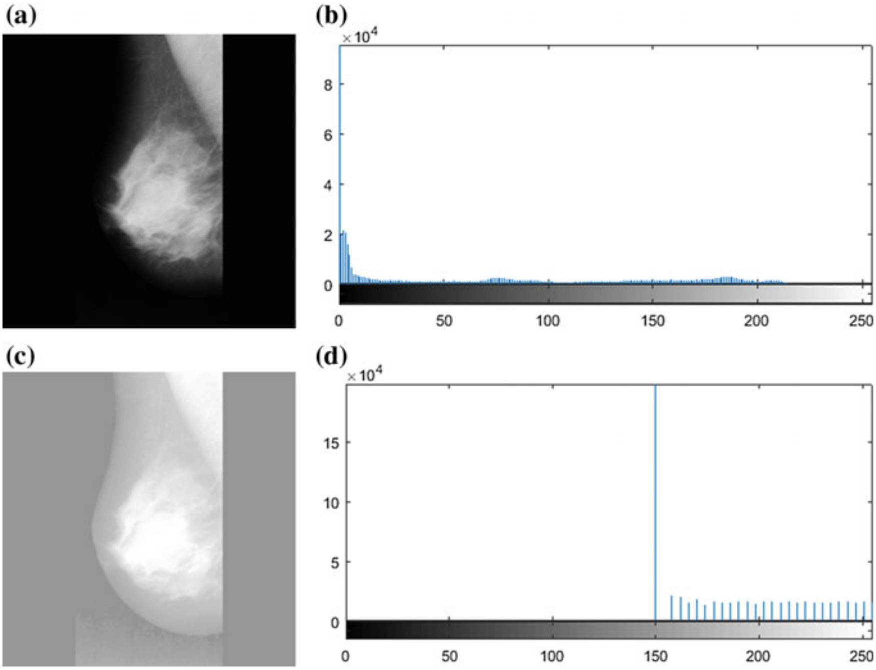


Fig. 3 Histogram equalization (HE) http://peipa.essex.ac.uk/pix/mias/all-mias.tar.gz

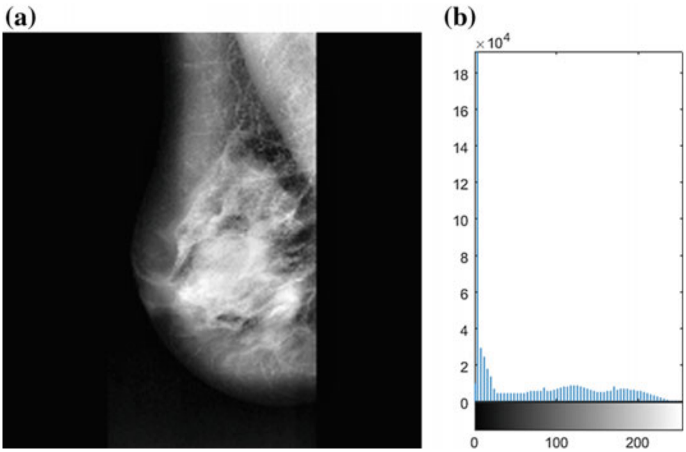


Fig. 4 Adjusted histogram equalization (AHE) http://peipa.essex.ac.uk/pix/mias/all-mias.tar.gz

image, (c) is the contrast adjusted image using Histogram Equalization, and (d) is the histogram of the contrast adjusted image.

2.2.2 Adjusted Histogram Equalization (AHE)

Figure 4a shows the contrast adjusted image and Fig. 4b shows the histogram obtained using adaptive histogram equalization "adapthisteq" applied to the image "mdb001.pgm" from the mini-MIAS database.

2.2.3 Contrast-Limited Adaptive Histogram Equalization

Contrast-Limited Adaptive Histogram Equalization (CLAHE) is applied to the image "mdb001.pgm" obtained from mini-MIAS database and the results are displayed in Fig. 5.

2.3 Spatial Filtering

Filtering is done to smoothen the image or upgrade or distinguish the edges in it. There are two filtering techniques:

- Linear
- Nonlinear

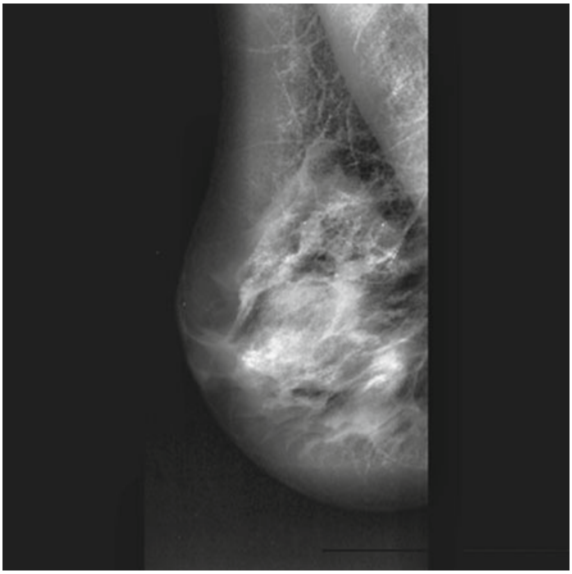


Fig. 5 Contrast-limited adaptive histogram equalization http://peipa.essex.ac.uk/pix/mias/all-mias.tar.gz

2.3.1 Linear

In a linear filter, the output will change linearly with a change in the input.
For Example,

(i) Gaussian.

Gaussian

Gaussian filtering is used to blur images and remove noise and detail. Figure 6a is the original image, (b) is the image with Gaussian noise, and (c) is the filtered image.

2.3.2 Nonlinear

There are various kinds of nonlinear filtering techniques such as

1. Mean filter
2. Median filter
3. Wiener filter.

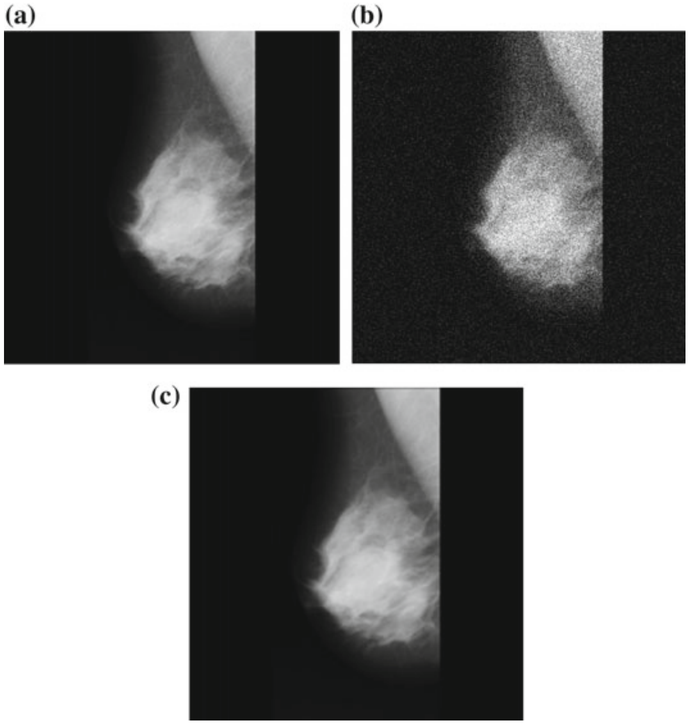


Fig. 6 Gaussian filter http://peipa.essex.ac.uk/pix/mias/all-mias.tar.gz

Mean Filter

Data set from mini-MIAS database has been obtained through PEIPA. In mean filter, at each position, the average value is assigned to the place of the center pixel. Mean filter was applied using MATLAB to the image ("mdb001.pgm") obtained from mini-MIAS database and the result is displayed in Fig. 7. Figure 7a is the original image, (b) is the image with salt-and-pepper noise introduced, and (c) is the filtered image.

Median Filter

Median filter is used to remove salt-and-pepper noise from the image using a nonlinear filtering technique. But Median filter cannot differentiate fine detail from noise. Figure 8a is the original image obtained from the mini-MIAS database, (b) is an image with salt-and-pepper noise. After applying median filter, the resulting image is displayed in (c).

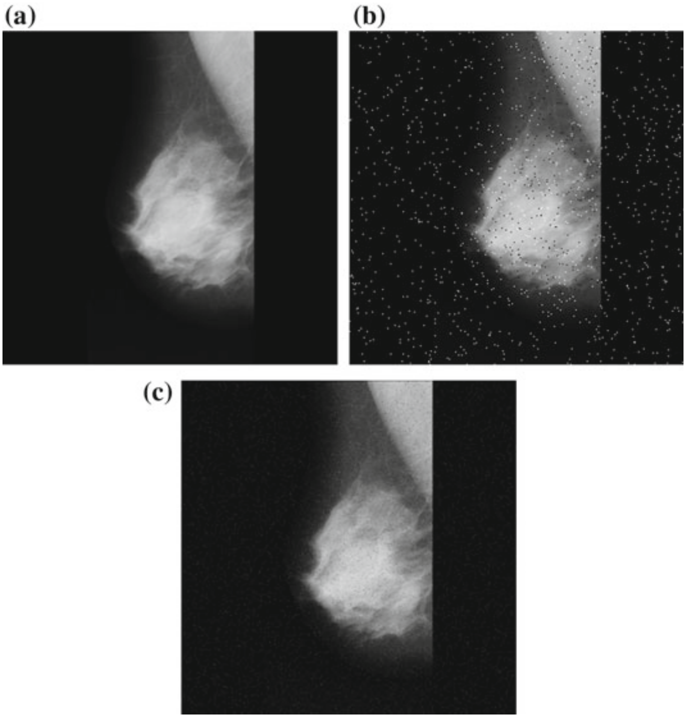


Fig. 7 Mean filter http://peipa.essex.ac.uk/pix/mias/all-mias.tar.gz

Wiener Filter

Wiener filter is used to remove the additive noise while inverting the blurring simultaneously. Figure 9a shows the original image taken from the mini-MIAS database, (b) is the image after introducing salt-and-pepper noise, (c) shows the filtered image.

3 Performance Analysis

3.1 Mean Square Error (MSE)

MSE of a filter is one of the ways to evaluate the difference between true value and the value implied by a filter. The difference exists as a consequence of randomness or because the filters may not consider information that could bring about a more veracious estimate. It corresponds to the expected value of square error loss. The average of squares of the error is gaged by MSE.

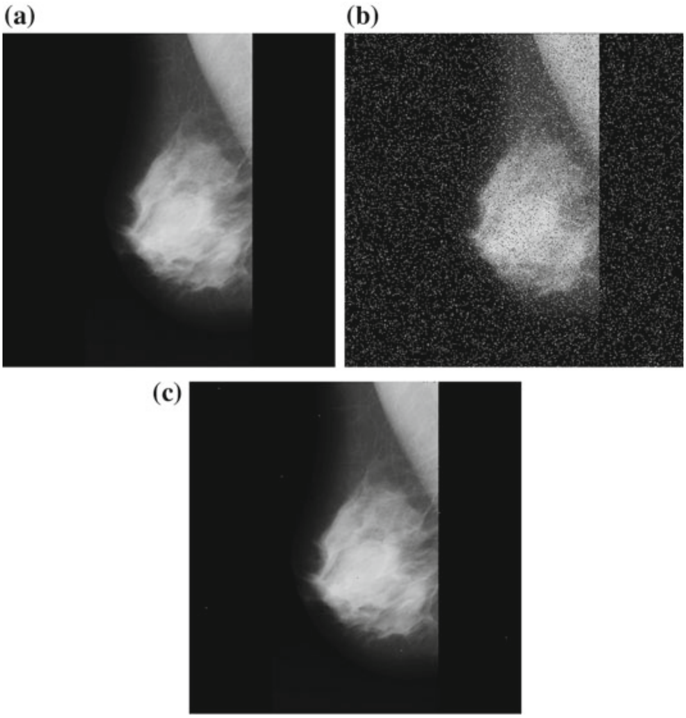


Fig. 8 Mean filter http://peipa.essex.ac.uk/pix/mias/all-mias.tar.gz

3.2 Peak Signal-to-Noise Ratio (PSNR)

PSNR is expressed in decibels and pixels in the image are assumed to be independent of their neighbors. Noise can get introduced to an original image due to compression, and PSNR is used to determine the image quality. For identical images, MSE is zero and PSNR becomes infinite. Higher value of PSNR implies that the image is of higher quality.

3.3 Contrast-to-Noise Ratio (CNR)

The quality of an image can be assessed using the CNR parameter. By improving CNR, the difference between two regions of interest is increased. Thus, the CNR is a variable index for evaluating detectability in digital radiography. Low-contrast lesion detection can be done using CNR between lesion and background. If the CNR is high, the lesion can be detected with higher probability.

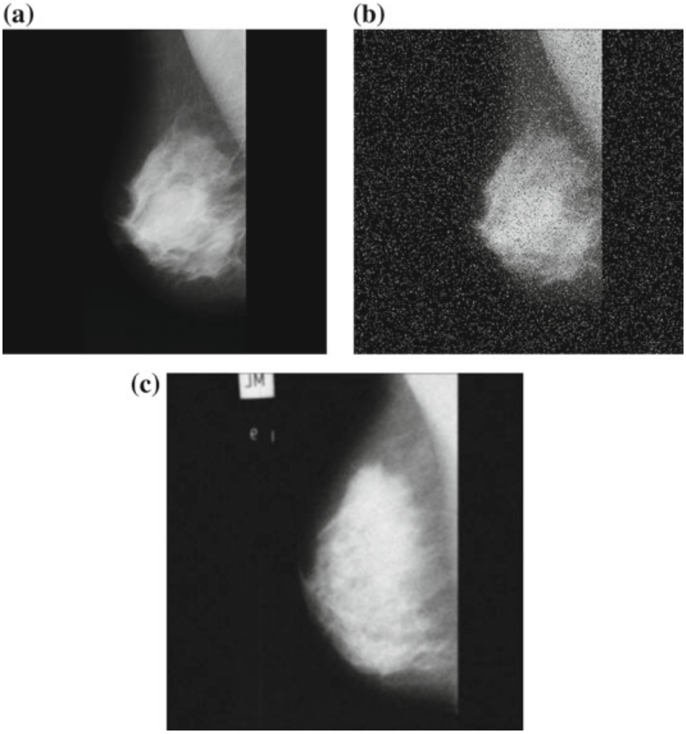


Fig. 9 Wiener filter http://peipa.essex.ac.uk/pix/mias/all-mias.tar.gz

4 Results and Discussion

Table 1 shows the values of MSE, PSNR, and CNR values of filters and image enhancement techniques applied. By comparing the performance of the filters, Gaussian filter has low MSE of 1.3935 and high PSNR of 36.6896. Among the image enhancement techniques used, contrast stretching has an MSE value of 2.8426 and PSNR value as 23.5935.

5 Conclusion

Hence, the preprocessing methods are discussed in this paper and Gaussian filter and contrast stretching are found to perform better when compared using MSE, PSNR, and CNR parameters with values 1.3935, 36.6896, and 6.8582 for Gaussian and 2.8426, 23.5935, and 1.6703 for contrast stretching, respectively.

Table 1 The values of MSE, PSNR, and CNR of filters and image enhancement techniques are applied

S.No	Technique	MSE	PSNR	CNR
1	Wiener filter	1.4194	26.9395	6.9742
2	Mean filter	5.8467	42.3309	6.7085
3	Average filter	4.2955	21.8138	1.4872
4	Median filter	4.6552	42.6932	1.2423
5	Rank filter	3.7373	2.416	1.7852
6	Gaussian filter	1.3935	36.6896	6.8582
7	Bit-plane slicing	2.1462	6.6833	6.9352
8	Histogram equalization (HE)	1.9396	5.253	1.3077
10	Adaptive HE (AHE)	3.8396	22.2879	2.7456
11	Contrast limited (AHE)	9.0199	18.5787	2.9872
12	Contrast stretching	2.8426	23.5935	1.6703

Compliance to Ethical Standards **Conflict of Interest**
Author A. R. Mrunalini, Author J. Premaladha declares that they have no conflict of interest.

Funding
We the authors would like to thank the Department of Science and Technology, India for their financial support through Fund for Improvement of S&T Infrastructure (FIST) programme (SR/FST/ETI-349/2013).

Ethical approval
This article does not contain any studies with human participants or animals performed by any of the authors.

Acknowledgement
We the authors sincerely thank the SASTRA Deemed to be University for providing an excellent infrastructure to carry out the research work.

References

1. Patel BK, Ranjbar S, Wu T, Pockaj BA, Li J, Zhang N, Lobbes M, Zhang B, Mitchell JR (2018) Computer-aided diagnosis of contrast-enhanced spectral mammography: a feasibility study. Eur J Radiol 31(98):207–213
2. Singh B, Kaur M (2018) An approach for classification of malignant and benign microcalcification clusters. Sādhanā 43(3):39
3. Khan KB, Khaliq AA, Jalil A, Shahid M (2018) A robust technique based on VLM and Frangi filter for retinal vessel extraction and denoising. PLoS ONE 13(2):e0192203
4. Shastri AA, Tamrakar D, Ahuja K (2018) Density-wise two stage mammogram classification using texture exploiting descriptors. Expert Syst Appl 1(99):71–82

5. Salem MA, Atef A, Salah A, Shams M (2018) Recent survey on medical image segmentation. In: Computer vision: concepts, methodologies, tools, and applications: concepts, methodologies, tools, and applications 2:129
6. de Moor T, Rodriguez-Ruiz A, Mann R, Teuwen J (2018) Automated soft tissue lesion detection and segmentation in digital mammography using a u-net deep learning network. ArXiv preprint arXiv:1802.06865
7. Diniz JO, Diniz PH, Valente TL, Silva AC, de Paiva AC, Gattass M (2018) Detection of mass regions in mammograms by bilateral analysis adapted to breast density using similarity indexes and convolutional neural networks. Comput Methods Programs Biomed
8. George MJ, Sankar SP. Efficient preprocessing filters and mass segmentation techniques for mammogram images. In: 2017 IEEE international conference on circuits and systems (ICCS). IEEE pp 408–413

A Study on Preprocessing Techniques for Ultrasound Images of Carotid Artery

S. Mounica, S. Ramakrishnan and B. Thamotharan

Abstract Ultrasound imaging has been widely used in the diagnosis of atherosclerosis. To precisely diagnose the carotid plaque, the affected region should be segmented from the ultrasonic image of carotid artery. Many techniques have been used to identify the plaque in ultrasound images. Image enhancement and restoration are the important processes to acquire high-quality images from the noisy images. When the artery images are captured, noise occurs due to high-frequency rate. To acquire a high-quality image, preprocessing is the first step to be done. The quality of the image is improved in this process. The techniques involved in preprocessing are dealt in this paper. Preprocessing involves filtering the image and removing the noise by various filtering techniques. Salt-and-pepper and Gaussian noise in ultrasound images can be filtered using techniques like mean, median and Wiener filters. Salt-and-pepper noise is multiplicative in nature and it is introduced by the image acquisition mechanism. The quality of the input sensor is reflected by the Gaussian noise. In this paper, the performance of image enhancement techniques on ultrasound images are evaluated using quality metrics, namely Mean Square Error (MSE) and Peak Signal–to-Noise Ratio (PSNR).

S. Mounica · S. Ramakrishnan · B. Thamotharan (✉)
Computer Vision & Machine Learning Laboratory, School of Computing, SASTRA Deemed University, Thanjavur, India
e-mail: balakrishthamo@gmail.com

S. Mounica
e-mail: sgammumounika@gmail.com

S. Ramakrishnan
e-mail: srk@ict.sastra.edu

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_159

1 Introduction

1.1 Ultrasound Carotid

Ultrasound imaging is a non-invasive procedure which uses sound waves to capture images that helps physicians to identify and treat illnesses conditions. Carotid artery is located on both sides of the neck which carries blood from the heart to the brain. Ultrasound image of the carotid arteries provides a clear understanding of these blood vessels and information about the blood flowing through them. The ultrasound imaging of carotid arteries in the neck is safe and painless. A Doppler ultrasound is one of the techniques that evaluates the blood flow through a blood vessel. A Doppler ultrasound study is a presurgical study used for examination of carotid artries. It is used to screen patients for Stenosis, a condition which increases the risk of stroke. It is also used for screening atherosclerosis which is caused by accumulation of lipids in carotid artries. Sample ultrasound image of the carotid artery is given in Fig. 1.

1.2 Preprocessing

The preprocessing of ultrasound images is done for removing speckle noise and noise due to wave interferences. Better segmentaiton can be achieved only after removing the above noises from the ultrasound image. Average and Bilateral filters not only

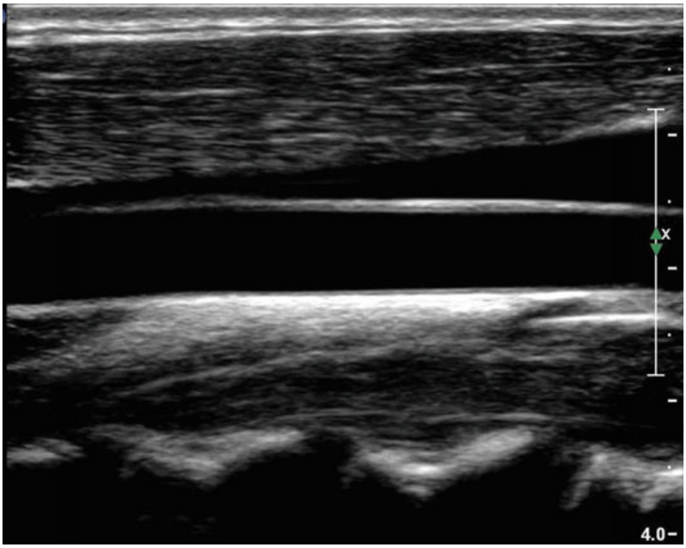


Fig. 1 Ultrasound image of carotid artery http://splab.cz/en/download/databaze/ultrasound

removes noise but also preserves the details of the ultrasound images of carotid artery [1].

1.3 Image Enhancement

Image enhacement is required to improve the quality of the input image. The image quality has some essential factors such as contrast, brightness, spatial resolution and noise. Image enhancement techniques are used for improving the above said quality factors of a image. A histogram gives the summary of the distribution of grey levels in an image by plotting the frequency of occurrence of grey levels. This plot provides a global distribution of appearance of the image that is used for image quality assessment. The image quality metrics are divided into objective and subjective fidelity criteria. The objective fidelity criteria is used to quantify the error that characterizes image quality. Subjective fidelity criteria depends on perception of the human observers and their visual systems.

1.4 Related Work

Datasets
The signal processing lab research group is a part of the Brno University of Technology and is involved in fields of voice, image, video and text processing which is used in fundamental and applied research. They have long-term experience in cooperation with many small, medium and supranational companies, and can offer advanced technologies and customized solutions in a variety of industrial fields. The biomedical signal processing focuses on signal analysis, which leads to the facilitation or improvement of diagnostic procedures and a better description or modelling of the human body as a complex dynamic spatial object. The key signals to be processed are image data produced by different instruments which are often in the form of temporal or spatial sequences (cuts). One-dimensional signals produced by different sensors are also used.

References	Techniques	Result
Chen et al. [2]	1. Butterworth filtering low pass and high pass 2. Adaptive weighted median filtering	Median filter removes the noise in the ultrasound image and helps to retain more information
Loizou et al. [3]	Median filter, Geometric filter, homomorphic filter, nonlinear coherence diffusion, Wavelet filter	The best performance is achieved by first-order statistics filter lmsv, geometric filter gf4d, homogeneous mask area filter
Shruthi et al. [4]	Median filter, Wiener filter, Gaussian low-pass filter, wavelet filter	Wavelet-based thresholding method gives better results in speckle noise reduction
Loizou et al. [5]	Normalization and speckle reduction filtering, a mean and variance of a pixel is utilized by Linear Scaling Mean Variance (LSMV)	Normalization and speckle reduction filtering give high quality image that can be used for accurate segmentation.
Abd-Elmoniem et al. [6]	NCD model, AWMF model	The article presents models that assists the segmentation technique and also area/volume calculation methods
Latha et al. [7]	Kaun filter Gabor filter Wavelet transform Adaptive filter	Kaun filter gives better results than average and bilateral filtering techniques
Noble et al. [8]	Speckle, K-distribution, Rician inverse of Gaussian distribution	The paper reviews ultrasound segmentation methods and highlighted the segmentation methods developed for medical B-mode ultrasound images.
Jeyalakshmi et al. [9]	Morphological Image Cleaning algorithm (MIC)	MMIC algorithm performs better in removing speckle noise and also preserves the features of the image.
Sudha et al. [10]	Wavelet domain noise filtering—Discrete Wavelet Transform (DWT)	When compared to other techniques, Wavelet based noise filtering provides image with improved visual quality and high Signal-to-Noise Ratio
Tang et al. [11]	ORIGIN, ENN, TOMEK	Filtering algorithm proposed in this article performs better than the other state of the art filtering techniques

2 Image Enhancement Techniques

2.1 *Spatial Domain*

In this domain, the point transforms or grey level scaling transformations to create the corresponding pixels in the out-take image depends only on the pixels of the intake image.

2.1.1 Linear Point Transformations

Inversion is one of the important transformations which performs a digital negative operation. In a binary image, inverse transformation reverses the image by changing a black pixel to a white one and vice versa.

2.1.2 Nonlinear Transformations

The relationship between both the input and output variables is not linear. So, to enhance the contrast of the given image, squared function is used and to compress the dynamic range of images, logarithmic function and exponential function are used to enhance the details in high-value regions of image and decrease the dynamic range in low-value regions, and power transformation is used to calculate the c which is positive constants.

2.2 *Histogram Processing*

Histogram techniques are effective and useful in image processing applications. It is used in manipulating the contrast and brightness of an image. It can be visualized as an intensity distribution or probability density function.

2.3 *Spatial Filtering*

Spatial filtering is the term which defines the filtering operations that are performed directly on the pixels of an image. It is used in image smoothing, sharpening, and noise removal. A filtered image is generated as the centre of the mask moves over every pixel in the input image.

2.3.1 Linear Filter

It is used to remove certain types of noise. If the out-put is a weighted sum of the input pixels, then the filtering method is linear. It works best for Gaussian noise. It is simple to execute and blurs the sharp edges. It is used to remove the lines and other details of the image. Gaussian and mean filters are commonly used linear filters.

2.3.2 Non-linear Filters

It is used to preserve edges and is an effective way to discard impulse noise. When compared to linear filters, non-linear filters provide better results. Non-linear filters remove noises without blurring the edges of the image. Median filter is the commonly used non-linear filter.

2.3.3 Gaussian Filter

Gaussian noise occurs when incidental values are added to an image due to incidental variations in the signal. This noise has normal distribution of the Probability Density Function (PDF). In digital images, it arises during acquisition and transmission. The techniques used to remove Gaussian noise are mean (convolutional), median and Gaussian filters (Fig. 2).

2.3.4 Salt-and-Pepper Noise

It is also called as impulse noise. The impulse noise is caused due to a pointed and abrupt disruption. Image with impulse noise will possess sparse black and white pixels as shown in Fig. 3.

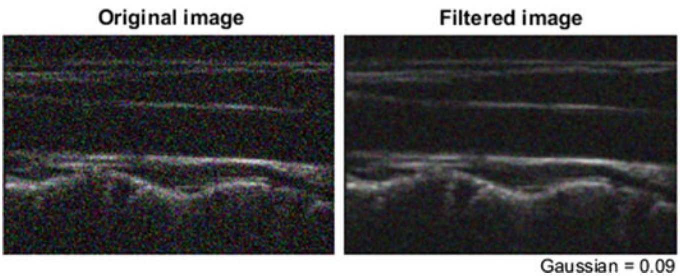


Fig. 2 Gaussian filter http://splab.cz/en/download/databaze/ultrasound

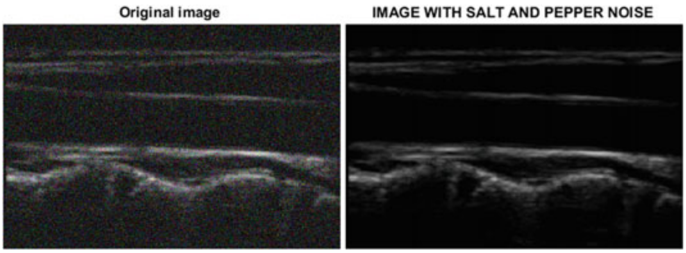


Fig. 3 Salt-and-pepper noise http://splab.cz/en/download/databaze/ultrasound

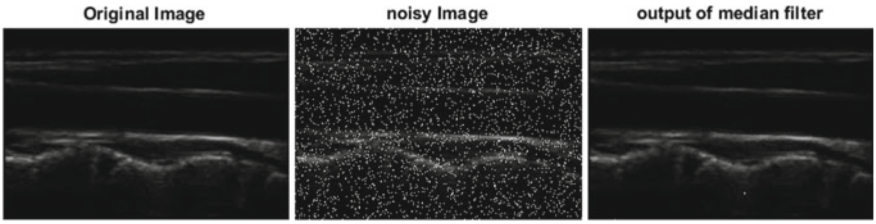


Fig. 4 Median filter http://splab.cz/en/download/databaze/ultrasound

2.3.5 Speckle Noise

The incidental values can be modelled by multiplying the pixel values of an image. Incidental fluctuation is a result that returns signal from an object. It increases the local area level of a mean grey. The filtering techniques used to remove speckle noise are mean and median filtering.

2.3.6 Median Filter

To preserve a useful detail of an image, median filter is often used than the mean filter. It is a nonlinear filtering technique that is used to remove noise from an image or signal by reducing the amount of intensity variation between one pixel and the other pixel. The median filter allows a high spatial frequency to pass through whereas remaining is used in removing noise on images. The affected one is less than half of the pixels in a smoothing neighbourhood [12]. When an image is corrupted by Gaussian noise, median filter is less effective at removing noise. Also, median filter is is relatively complex to compute and expensive. (Fig. 4).

2.3.7 Mean Filter

Mean filtering often uses averaging kernel. It is a simple linear filter that replaces each pixel value in an image with the mean value of its neighbourhood to remove the impulse noise. Mean filter is often used for smoothing. Nath et al. says that the convolution filter is used in mean filtering. Like other convolutions, the shape and size of the neighbourhood to be sampled is based on the kernel while calculating the mean [13] (Fig. 5).

2.3.8 Wiener Filter

Weiner filter remove the noise that corrupts the image. It is based on statistical approach. Reducing the mean square error is the main goal [12].

The spectral property of noise and the original signal is based on knowledge of

$$G(u, v) = H * (u, v) \frac{H * (u, v)}{|H(u, v)|^2 Ps(u, v) + Pn(u, v)}$$

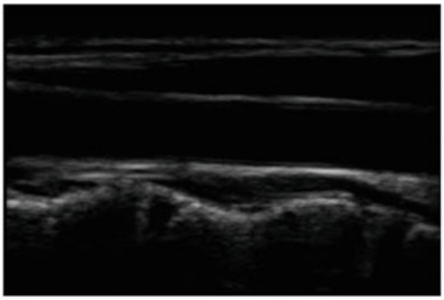
The Wiener filter Fourier domains are

$H^*(u, v)$ complex conjugate of degradation function,
$\mathrm{Pn}(u, v)$ power spectral density of noise,
$\mathrm{Ps}(u, v)$ power spectral density of non-degraded image (Fig. 6).

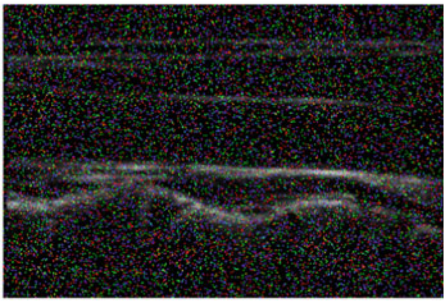
2.3.9 Fourier Transform

See Figs 7, 8.

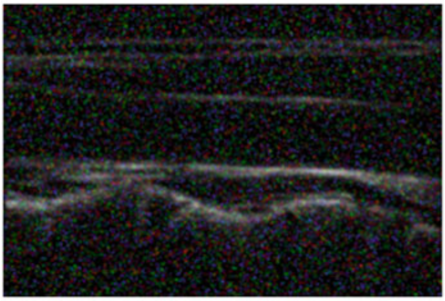
Fig. 5 Mean filter



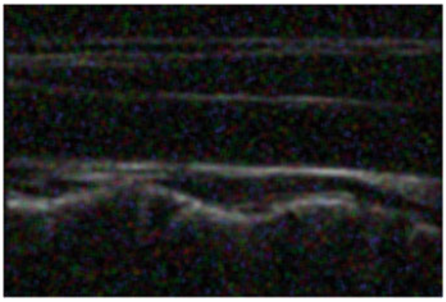
(a) original image(http://splab.cz/en/download/databaze/ultrasound)



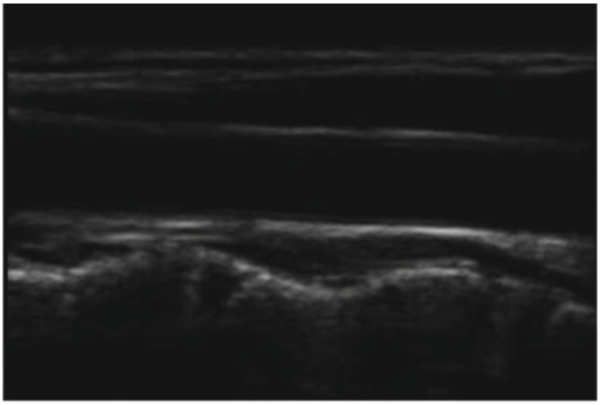
(b) Noisy image



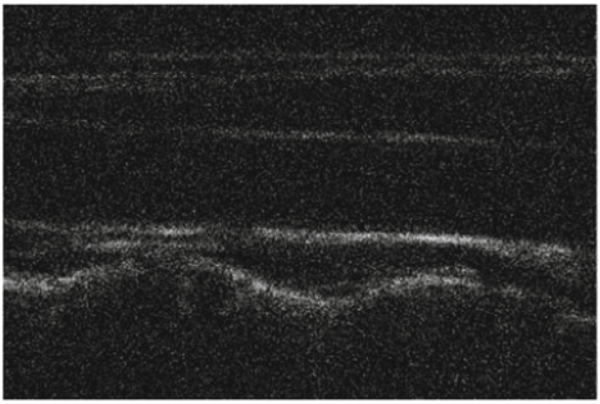
(c) A Mean filter



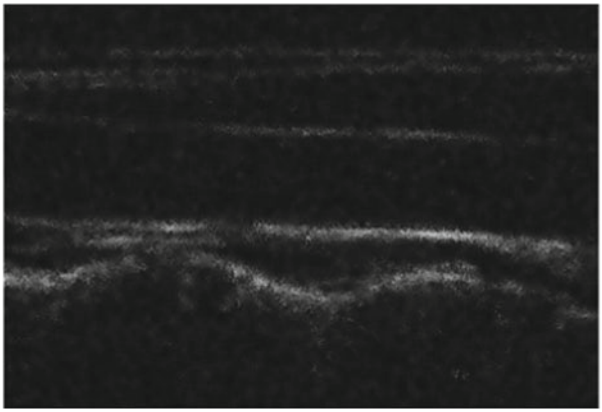
(d) B Mean filter



(a) original image(http://splab.cz/en/download/databaze/ultrasound)



(b) Noisy Image



(c) Weiner filter

Fig. 6 Weiner filter

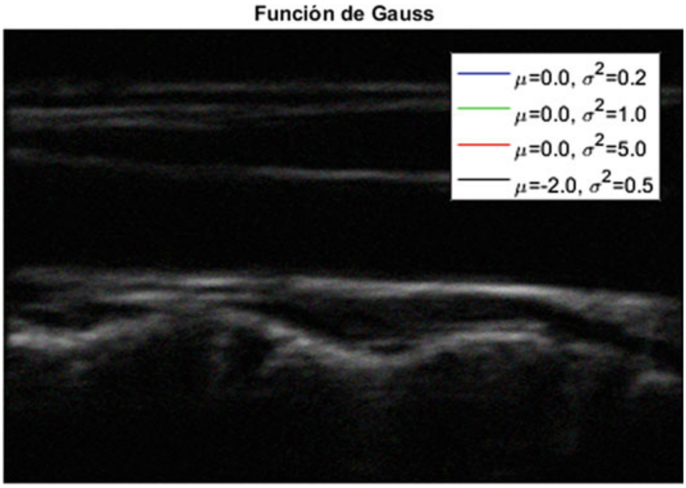


Fig. 7 Gaussian http://splab.cz/en/download/databaze/ultrasound

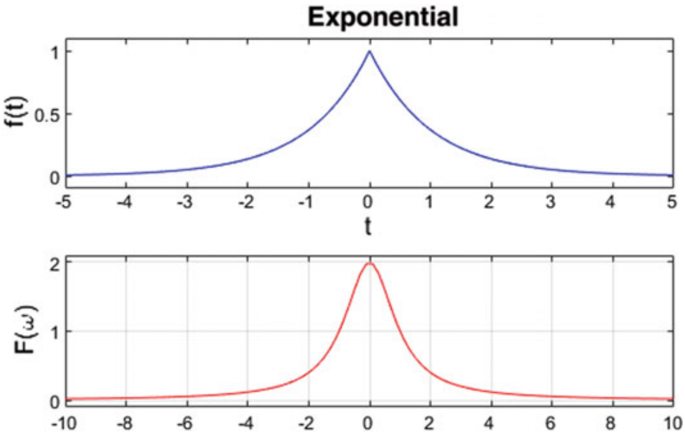


Fig. 8 Exponential

3 Performance Analysis

3.1 Mean Square Error (MSE)

Mean Square Error (MSE) of an image measures the average of squares of errors or deviations i.e., difference between original image and filtered image.

The MSE is measured between the encoded and the original image which is defined by a cumulative square error

$$\text{MSE} = \frac{1}{mn}\sum_{0}^{m-1}\sum_{0}^{n-1}\|f(i,j) - g(i,j)\|^2$$

3.2 Peak Signal-to-Noise Ratio (PSNR)

The ratio between the maximal available value (power) of a signal and the power of deceiving noise is termed as peak signal-to-noise ratio which affects the condition of its representation. It is usually expressed in logarithmic decibel scale. It is an approximation of reconstruction quality in human perception.

$$\text{PSNR} = 20\log_{10}\left(\frac{\text{MAX}_f}{\sqrt{\text{MSE}}}\right)$$

4 Results and Discussion

S.No	Technique	MSE	PSNR
1	Guassian filter	23.0816	17.6008
2	Mean filter	37.79	32.38
3	Median filter	33.17	32.96
4	Weiner filter	38.87	32.25
5	Salt-and-pepper noise	35.11	32.72
6	Fourier transform	36.20	32.60

5 Conclusion

The most necessary task in image processing is the enhancement of an input image. For removing the noise from the images, filters are required. In this paper, different filters are implemented for different types of noisy ultrasound images and it is concluded that median filter performs better. Higher the PSNR value, lower the MSE value, higher the quality of the image. Ultrasound image enhanced with median filter results in higher PSNR value and lower MSE value. Enhanced image can be used for further image processing steps like segmentation, feature extraction, etc.

Compliance to Ethical Standards Conflict of Interest
Author S. Mounica, Author S. Ramakrishnan and Author B. Thamotharan declares that they have no conflict of interest.

Funding
We the authors would like to thank the Department of Science and Technology, India for their financial support through Fund for Improvement of S&T Infrastructure (FIST) programme (SR/FST/ETI-349/2013).

Ethical approval
This article does not contain any studies with human participants or animals performed by any of the authors.

Acknowledgement
We the authors sincerely thank the SASTRA Deemed to be University for providing an excellent infrastructure to carry out the research work.

References

1. Kazubek M (2003) Wavelet domain image denoising by thresholding and wiener filtering. IEEE Signal Process Lett 10(11):324–326
2. Chen J, Li F, Fu Y, Liu Q, Huang J, Li K (2017) A study of image segmentation algorithms combined with different image preprocessing methods for thyroid ultrasound images. IEEE ICIST:1–5
3. Loizou CP, C.S. Pattichis, C.I. Christodoulou, R.S.H. Istepanian, M. Pantziaris, A. Nicolaides(2005) Comparative evaluation of despeckle filtering in ultrasound imaging of the carotid artery. IEEE Trans Ultrason Ferroelectr Freq Control 52 (10):1653–1669
4. Shruthi B, Renukalatha S, Siddappa DM (2015) Speckle noise reduction in Ultrasound images—A review. IJERT 4(02):1402–1406
5. Loizou CP, Pattichis CS, Pantziaris M, Tyllis T, Nicolaides A (2006) Quality evaluation of ultrasound imaging in the carotid artery based on normalization and speckle reduction filtering. Med Biol Eng Compu 44(5):414
6. Abd-Elmoniem KZ, A.-B.M. Youssef, Y.M. Kadah, (2002) Real-time speckle reduction and coherence enhancement in ultrasound imaging via nonlinear anisotropic diffusion. IEEE Trans Biomed Eng 49(9):997–1014
7. Latha S, Dhanalakshmi S, Muthu P (2016) A Review and Comparative Study of Methods used in Finding Carotid Artery Abnormalities using Ultrasound Images. IJCTA 9(10):4891–4898
8. Noble JA, Boukerroui D (2006) Ultrasound image segmentation: a survey. IEEE Trans Med Imaging 25(8):987–1010

9. Jeyalakshmi TR, Ramar K (2010) A modified method for speckle noise removal in ultrasound medical images. Int J Comput Electr Eng 2(1):54
10. Sudha S, Suresh GR, Sukanesh R (2009) Speckle Noise Reduction in Ultrasound Images by Wavelet Thresholding based on Weighted Variance. Int J of Comput Theory Eng 7–12
11. Tang S, Chen SP (2008) An effective data preprocessing mechanism of ultrasound image recognition. IEEE ICBBE 2708–2711
12. Kumar S, Kumar P, Gupta M, Nagawat AK (2010) Performance comparison of median and wiener filter in image de-noising. Int J Comput Appl 12:0975–8887
13. Nath A (2013) Image denoising algorithms: a comparative study of different filtration approaches used in image restoration. In: 2013 International conference on communication systems and network technologies (CSNT). IEEE, pp 157–163
14. Paini A, Boutouyrie P, Calvet D, Zidi M, Agabiti-Rosei E, Laurent S Multiaxial mechanical characteristics of carotid plaque: analysis by Multiar

Fractional Reaction Diffusion Model for Parkinson's Disease

Hardik Joshi and Brajesh Kumar Jha

Abstract Calcium (Ca^{2+}) ion known as a second messenger, involve in variety of signalling process, and directly link with the intracellular calcium concentration ($[Ca^{2+}]$) that are continuously remodelled for the survival of the nerve cell. Buffer, also refer as a protein, react with Ca^{2+} and significantly lower down the intracellular $[Ca^{2+}]$ in nerve cell. There are numerous signalling processes in mammalian brain which can initiate at the high level of intracellular $[Ca^{2+}]$. Voltage gated calcium channel (VGCC), and ryanodine receptor (RyR) are work as an outward source of Ca^{2+} which initiate, and sustain the signalling process for smooth functioning of the cells. Parkinson's disease (PD) is a brain disorder of the central nervous system accompanied with the alteration of the signalling process. In present paper, a one dimensional fractional reaction diffusion model is consider to understand the physiological role of buffer, VGCC, and RyR in view of the PD.

1 Introduction

Parkinson's disease (PD) is second most common neurodegenerative disorder spread in the world after the Alzheimer's disease [1]. The movement symtoms of PD is tremor, rigidity, bradykinesia, and postural instability caused by the loss or dysfunction of the neurotransmitter dopamine in the midbrain [1]. The alteration in the signalling process resultant transient changes in spatial as well as the temporal intracellular concentration. The high level of intracellular concentration is responsible for the apoptosis or the dysfunction of nerve cells [2] which produce the neurotransmitter dopamine [3]. Ca^{2+} being a second messenger play a vital role in variety of physiological function which is essential for the signalling process, like movement

H. Joshi (✉) · B. K. Jha
Department of Mathematics, School of Technology,
Pandit Deendayal Petroleum University, Gujarat, India
e-mail: hardik.joshi8185@gmail.com

B. K. Jha
e-mail: brajeshjha2881@gmail.com

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_160

of the muscle, fertilization, proliferation, cellular motility, memory of the cell, etc. [4]. Ca^{2+} make a place making role in the presence of intracellular and extracellular source of Ca^{2+}. Buffer is one of the intracellular source of Ca^{2+}. Calbindin-D_{28k} is a Ca^{2+} binding buffer, they react with the free Ca^{2+} in the cytosol, and make a Ca^{2+} bound buffer [5]. The overexpression of Ca^{2+} bound buffer helps to decrease the peak of intracellular $[Ca^{2+}]$ in the nerve cell [6–8]. But there are various process inside the cell which can be initiate at the threshold level or at the presence of high intracellular $[Ca^{2+}]$. Voltage gated calcium channel (VGCC) and ryanodine receptor (RyR) are the extracellular source of Ca^{2+} which are normally closed at resting position of the cells. The opening of VGCC and RyR initiate as well as sustain many of physiological function of the nerve cells [8]. The alteration or dysfunction of these parameters may lead to resultant transient changes in spatially as well as temporally, which is the pathological symptoms of PD [2].

Literature survey evidently show that, the effect of buffer on Ca^{2+} distribution have been reported on the different cells like neuron, astrocytes, oocytes, fibroblast, myocytes, etc. [4, 7–12]. Tewari and Pardasani have studied the cytosolic Ca^{2+} diffusion in presence of excess buffer [9]. Jha et al. have studied Ca^{2+} distribution in presence of excess buffer on astrocytes cell [4]. Panday and Pardasani have studied Ca^{2+} distribution in oocytes cell [10]. In the same year, Kotwani et al. have studied Ca^{2+} distribution in fibroblast cell [11]. Pathak and Adlakha have studied Ca^{2+} distribution in cardiac myocytes [12]. All the authors [4, 8–12] have employed the finite element method to obtain the numerical results. In recent year, Dave and Jha have studied the effect of buffer on Ca^{2+} diffusion in view of neurodegenerative disease. They have adopted two dimensional advection reaction equation and applied Laplace and similarity transformation to obtain the analytical solutions. Finally, they interpreted the obtained results with the physiology of Alzheimer's disease [13].

Here an attempt has been made to study the one dimensional fractional reaction diffusion model in presence of calbindin-D_{28k}, VGCC, and RyR. The obtained analytical results are stimulate in the MATLAB, and incorporate with the physiology of the Parkinson's brain. The main motive to study reaction diffusion model by applying fractional approach is, due to its non-local property. The integer order ordinary or partial differential equation has a local property. It means that to obtain the results at any stage of the entire process, is considered the input at that particular stage. On the other hand in fractional differential equation it compute the results at any stage of the process by taking all the input from the resting stage of the process [14, 15]. Ca^{2+} diffusion in the nerve cell is a complex dynamics process occurs in mammalian brain, and calculated in μs, so in view of this reaction diffusion model is studied fractionally.

2 Mathematical Model

Ca^{2+} kinetics in the human nerve cells is governed by a set of reaction diffusion equations which can be given by the bimolecular reaction between Ca^{2+} and buffer species [4, 7, 16]

$$\left[\mathrm{Ca}^{2+}\right] + [\mathrm{B}] \underset{k^-}{\overset{k^+}{\Leftrightarrow}} [\mathrm{CaB}] \tag{1}$$

where $[Ca^{2+}]$, [B], and [CaB] are intracellular $[Ca^{2+}]$, free and bound buffers respectively.

$$\frac{\partial\left[\mathrm{Ca}^{2+}\right]}{\partial t} = D_{\mathrm{Ca}}\nabla^2\left[\mathrm{Ca}^{2+}\right] + \sum_j R_j \tag{2}$$

$$\frac{\partial[B]}{\partial t} = D_B\nabla^2[B] + R_j \tag{3}$$

$$\frac{\partial[\mathrm{CaB}]}{\partial t} = D_{\mathrm{CaB}}\nabla^2[\mathrm{CaB}] - R_j \tag{4}$$

where

$$R_j = -k^+[\mathrm{B}]\left[\mathrm{Ca}^{2+}\right] + k^-[\mathrm{CaB}] \tag{5}$$

D_{Ca}, D_B, D_{CaB} are diffusion coefficients of free Ca^{2+}, free buffer and Ca^{2+} bound buffers respectively. k^+ and k^- are association and dissociation rate constants for buffer 'j' respectively. Then by combining Eqs. (1–5) the mathematical model can be frame as

$$\frac{\partial^\alpha[\mathrm{Ca}^{2+}]}{\partial t^\alpha} = D_{\mathrm{ca}}\frac{\partial^\beta[\mathrm{Ca}^{2+}]}{\partial x^\beta} - k_j^+[B]_\infty([\mathrm{Ca}^{2+}] - [\mathrm{Ca}^{2+}]_\infty) + J_{\mathrm{VGCC}} + J_{\mathrm{RyR}} \tag{6}$$

where $\frac{\partial^\alpha[\mathrm{Ca}^{2+}]}{\partial t^\alpha}$ and $\frac{\partial^\beta[\mathrm{Ca}^{2+}]}{\partial x^\beta}$ is the Caputo time fractional derivative of order $\alpha(0 < \alpha \leq 1)$ and Caputo space fractional derivative of order $\beta(1 < \beta \leq 2)$ respectively [17].

J_{VGCC} is the Ca^{2+} influx through VGCC and modelled using the Goldman-Hodgkin-Katz (GHK) current equation given by Jha et al. [6]

$$I_{\mathrm{Ca}} = P_{\mathrm{Ca}} z_{\mathrm{Ca}}^2 \frac{F^2 V_m}{RT} \frac{[\mathrm{Ca}^{2+}]_i - [\mathrm{Ca}^{2+}]_o \exp(-z_{\mathrm{Ca}}\frac{FV_m}{RT})}{1 - \exp(-z_{\mathrm{Ca}}\frac{FV_m}{RT})} \tag{7}$$

where $[Ca^{2+}]_i$ and $[Ca^{2+}]_o$ are the intracellular and extracellular $[Ca^{2+}]$ respectively; P_{Ca} is the permeability of Ca^{2+} ion; z_{Ca} is the valance of Ca^{2+} ion; F is Faraday's constant; V_m is the membrane potential; R is the real gas constant; T is the absolute temperature. The current equation is converted into molar/second by using [6]

$$\sigma_{\text{Ca}} = -\frac{I_{\text{Ca}}}{z_{\text{Ca}} F V_{\text{Nervecells}}} \tag{8}$$

where negative sign indicated that the inward current is negative.

J_{RyR} is the Ca^{2+} influx through RyR given as [8]

$$J_{\text{RyR}} = V_{\text{RyR}} P_o([\text{Ca}^{2+}]_{\text{ER}} - [\text{Ca}^{2+}]) \tag{9}$$

Combining Eqs. (6–9), the proposed mathematical model converted as,

$$\begin{aligned}\frac{\partial^\alpha [\text{Ca}^{2+}]}{\partial t^\alpha} = D_{\text{ca}} \frac{\partial^\beta [\text{Ca}^{2+}]}{\partial x^\beta} - k_j^+ [B]_\infty([\text{Ca}^{2+}] - [\text{Ca}^{2+}]_\infty) \\ + P_{\text{Ca}} z_{\text{Ca}}^2 \frac{F^2 V_m}{RT} \frac{[\text{Ca}^{2+}]_i - [\text{Ca}^{2+}]_o \exp\left(-z_{\text{Ca}} \frac{F V_m}{RT}\right)}{1 - \exp\left(-z_{\text{Ca}} \frac{F V_m}{RT}\right)} \\ + V_{\text{RyR}} P_o([\text{Ca}^{2+}]_{\text{ER}} - [\text{Ca}^{2+}])\end{aligned} \tag{10}$$

Along with initial and boundary conditions as,

$$[\text{Ca}^{2+}](x, 0) = g(x), \ [\text{Ca}^{2+}](\pm\infty, t) = 0 \tag{11}$$

The Eq. (10) can be rewritten as

$$\frac{\partial^\alpha C}{\partial t^\alpha} = D_{\text{ca}} \frac{\partial^\beta C}{\partial x^\beta} - aC + b \tag{12}$$

where $C = [\text{Ca}^{2+}]$,

$$a = k_j^+ [B]_\infty + \frac{P_{\text{Ca}} z_{\text{Ca}} F V_m}{RT V_{\text{Nervecells}}} \frac{\exp\left(z_{\text{Ca}} \frac{F V_m}{RT}\right)}{1 - \exp\left(z_{\text{Ca}} \frac{F V_m}{RT}\right)} + V_{\text{RyR}} P_o \tag{13}$$

and

$$b = k_j^+ [B]_\infty C_\infty + \frac{P_{\text{Ca}} z_{\text{Ca}} F V_m}{RT V_{\text{Nervecells}}} \frac{[\text{Ca}^{2+}]_o}{1 - \exp(z_{\text{Ca}} \frac{F V_m}{RT})} + V_{\text{RyR}} P_o C_{\text{ER}} \tag{14}$$

Applying temporal fractional Laplace transform and spatial fractional Fourier transform technique on Eq. (12) along with Eq. (11), the solution of Eq. (12) is obtained in terms of green function

$$G_{\alpha,\beta}(x,t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathrm{e}^{-ikx} E_\alpha\left[\left\{D_{\mathrm{Ca}}(-ik)^\beta - a\right\}\cdot t^\alpha\right]\mathrm{d}k + \frac{bt^\alpha}{2\pi}\int_{-\infty}^{\infty} \mathrm{e}^{-ikx} E_{\alpha,\alpha+1}\left[\left\{D_{\mathrm{Ca}}(-ik)^\beta - a\right\}\cdot t^\alpha\right]\mathrm{d}k \tag{15}$$

where

$$E_\alpha(z) = \sum_{k=0}^{\infty}\frac{z^k}{\Gamma(\alpha k+1)}, \quad \mathrm{R}(\alpha) > 0, \quad \alpha, z \in Ç \text{ and}$$

$$E_{\alpha,\beta}(z) = \sum_{k=0}^{\infty}\frac{z^k}{\Gamma(\alpha k+\beta)}, \quad \mathrm{R}(\alpha), R(\beta) > 0, \quad \alpha, \beta, z \in Ç$$

are the Mittag-Leffler function for one and two parameter respectively [17].

The solution of time fractional reaction diffusion equation can be obtained by varying fractional order α in Eq. (15), keeping fix space derivative $\beta = 2$. Again by applying fractional Laplace and Fourier transform technique, the solution corresponding to time fractional reaction diffusion equation is

$$G_{\alpha,2}(x,t) = \frac{1}{2\sqrt{\pi D_{\mathrm{Ca}} t^\alpha}}\int_0^\infty \mathrm{e}^{-\frac{x^2}{4D_{\mathrm{Ca}}t^\alpha k} - \mathrm{at}^\alpha k} k^{-\frac{1}{2}} M_\alpha(k)\mathrm{d}k + \frac{bt^\alpha}{2\sqrt{\pi D_{\mathrm{Ca}} t^\alpha}}\int_0^\infty \mathrm{e}^{-\frac{x^2}{4D_{\mathrm{Ca}}t^\alpha k} - \mathrm{at}^\alpha k} k^{-\frac{1}{2}} \phi(-\alpha, 1; -k)\mathrm{d}k \tag{16}$$

where

$$M_\alpha(z) = \sum_{k=0}^{\infty}\frac{(-1)^k z^k}{\Gamma(-\alpha k + (1-\alpha))\cdot k!}, \quad 0 < \alpha < 1 \text{ and}$$

$$\phi(\alpha, \beta; z) = \sum_{k=0}^{\infty}\frac{z^k}{\Gamma(\alpha k+\beta)\cdot k!}, \quad \alpha > -1, \quad \beta \in Ç$$

are the Mainardi function and Wright function respectively [17].

The solution of space fractional reaction diffusion equation can be obtained by varying fractional order β in Eq. (15), keeping fix time derivative $\alpha = 1$. By applying same fractional Fourier transform technique, we get the solution corresponding to space fractional reaction diffusion equation in terms of Levy stable distribution [18]

$$G_{1,\beta}(x,t) = \frac{e^{-at}}{\left(D_{Ca_\alpha}t\right)^{1/\beta}} S\left(\frac{x}{\left(D_{Ca_\alpha}t\right)^{1/\beta}}/\beta, 1, 1, 0; 1\right) + \frac{bt}{\left(D_{Ca_\alpha}t\right)^{1/\beta}\left\{\ln S\left(\frac{x}{\left(D_{Ca_\alpha}t\right)^{1/\beta}}/\beta, 1, 1, 0; 1\right) - at\right\}} \times \left[\left\{\frac{e^{-at}}{\left(D_{Ca_\alpha}t\right)^{1/\beta}} S\left(\frac{x}{\left(D_{Ca_\alpha}t\right)^{1/\beta}}/\beta, 1, 1, 0; 1\right)\right\} - 1\right] \tag{17}$$

3 Results and Discussion

The values of physiological parameters used to simulate the results are given in Table 1 or stated along with the figure. Figure 1 show the temporal distribution of $[Ca^{2+}]$ in presence of buffer, VGCC, and RyR. The value of the temporal and spatial order derivative in Fig. 1a–e is, in figure (a) $\alpha = 1$ and $\beta = 2$, (b) $\alpha = 1$ and $\beta = 1.9$, (c) $\alpha = 1$ and $\beta = 1.8$, (d) $\alpha = 0.9$ and $\beta = 2$, and (e) $\alpha = 0.8$ and $\beta = 2$ respectively. In Fig. 1, the value of $x = 0.001$ μm i.e. near the source of the channel. In all Fig. 1a–e, it can be observed that increase the amount of buffer concentration, lower down the profile of $[Ca^{2+}]$. Physiologically, it happens due to more amount of buffer react with free Ca^{2+} ion in the cell and make Ca^{2+} bound buffer to control the intracellular $[Ca^{2+}]$. The Fig. 1a is corresponding to $\alpha = 1$ and $\beta = 2$, it means

Table 1 Values of physiological parameters [5, 6, 8]

Symbol	Parameter	Value
D_{Ca}	Diffusion coefficient	250 $\mu m^2/s$
$[B]$	Buffer concentration	100–350 μM
$[Ca^{2+}]_\infty$	Background $[Ca^{2+}]$	0.1 μM
k^+	Buffer association rate	75 $\mu M^{-1} s^{-1}$
$V_{Nervecells}$	Volume of nerve cells	$5.233 * 10^{13}$ l
R	Gas constant	8.31 J/(mole · K)
T	Temperature	300 K
Z_{Ca}	Valence of Ca^{2+} ion	2
P_{Ca}	Permeability of Ca^{2+}	$4.3 * 10^{-8}$ m/s
V_m	Membrane potential	−0.05 V
V_{RyR}	RyR rate	0.5 μM/s
P_o	Rate of Ca^{2+} efflux	0.5 M/s
$[Ca^{2+}]_{ER}$	ER $[Ca^{2+}]$	500 μM

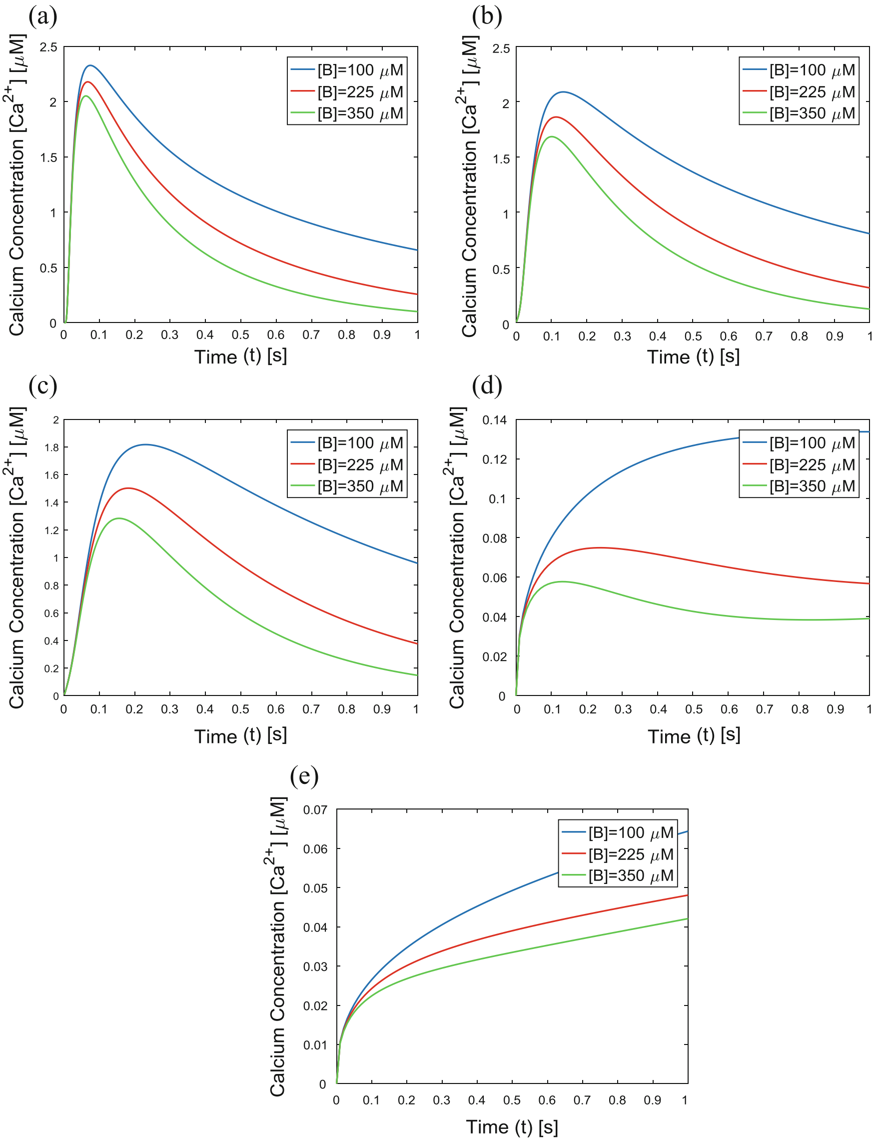


Fig. 1 Temporal distribution of [Ca^{2+}] in presence of Buffer, VGCC, and RyR with **a** $\alpha = 1$ and $\beta = 2$, **b** $\alpha = 1$ and $\beta = 1.9$, **c** $\alpha = 1$ and $\beta = 1.8$, **d** $\alpha = 0.9$ and $\beta = 2$, and **e** $\alpha = 0.8$ and $\beta = 2$

the standard diffusion model for Ca^{2+} dynamics, which can be easily validate for the previous obtained results [8]. The main advantage of this study is obtained the results of Ca^{2+} dynamics in nerve cells at any fraction of time, which can be easily visualized in Fig. 1b–e.

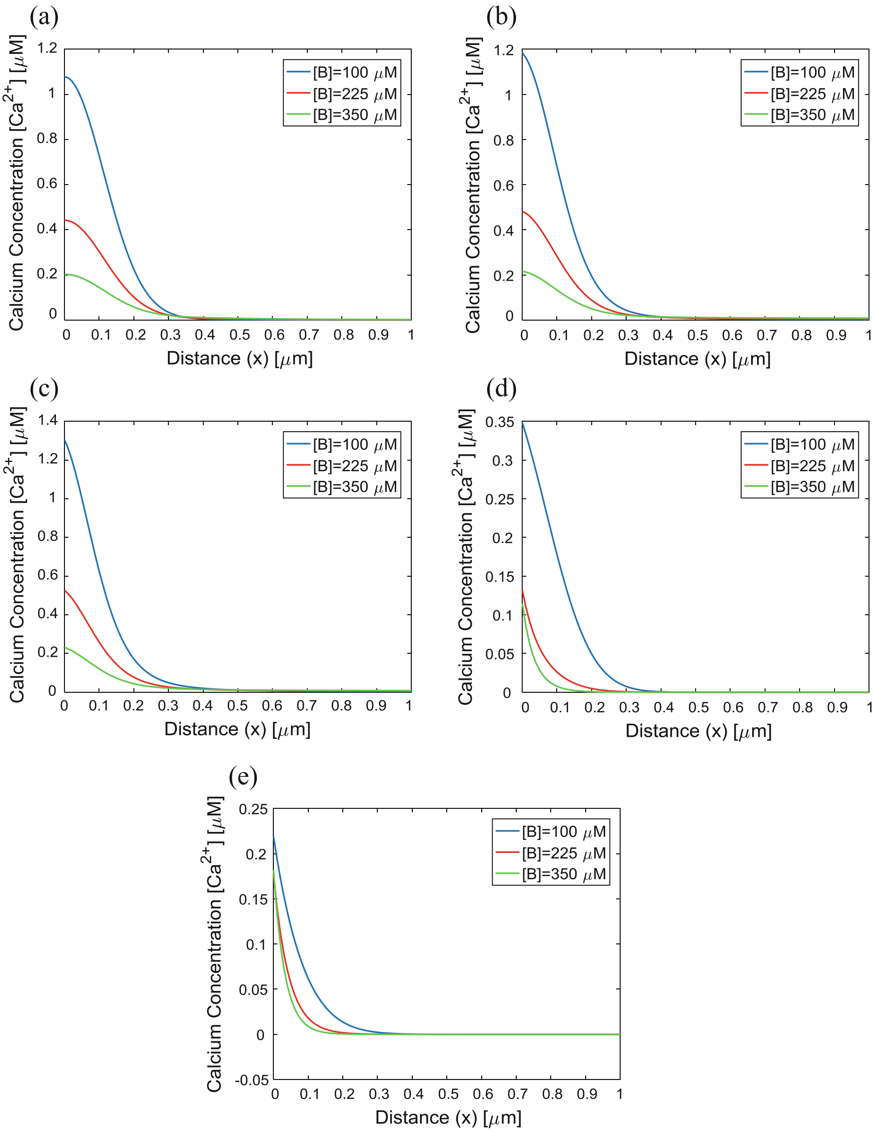


Fig. 2 Spatial distribution of $[Ca^{2+}]$ in presence of Buffer, VGCC, and RyR with **a** $\alpha = 1$ and $\beta = 2$, **b** $\alpha = 1$ and $\beta = 1.9$, **c** $\alpha = 1$ and $\beta = 1.8$, **d** $\alpha = 0.9$ and $\beta = 2$, and **e** $\alpha = 0.8$ and $\beta = 2$

Figure 2 show the spatial distribution of $[Ca^{2+}]$ in presence of buffer, VGCC, and RyR. The value of the temporal and spatial order derivative in Fig. 2a–e is same as in Fig. 1a–e. In Fig. 2, the value of $t = 0.01$ s. In all Fig. 2a–e, it can be observed that increase the amount of buffer concentration, reduce the profile of $[Ca^{2+}]$ due to the same physiologically as mention above. The Fig. 2a is corresponding to $\alpha = 1$

and $\beta = 2$, it means the standard spatial diffusion for Ca^{2+} dynamics, which can be easily validate the previously obtained results [8]. It is observed from the figure that for smaller value of fractional order the solution become to much closer and achieve the steady state as soon as possible.

4 Conclusion

The fractional model studied in this paper provide interesting results in presence of buffers, VGCC, and RyR in view of the progression of the disease. The buffer play a vital role in gradually reduced the amount of the intracellular $[Ca^{2+}]$. Whereas VGCC, and RyR have significant effect in the presence of low amount of calbindin-D_{28k}. There are complex interplay among the buffer, VGCC, and RyR for maintaining the adequate level of intracellular concentration for the nerve cells. The dysfunctioning or alteration in these complex interplay associated with the brain disorder, PD. The obtained results can be helps to mediate or slower down the progression of the Parkinson's.

Acknowledgements We are very grateful to the reviewers for their fruitful comments and suggestions for making updation in the research article.

References

1. Surmeier DJ, Schumacker PT, Guzman JD, Ilijic E, Yang B, Zampese E (2017) Calcium and parkinson's disease. Biochem Biophys Res Commun. 483:1013–1019
2. Jha BK, Joshi H, Dave DD (2016) Portraying the effect of calcium-binding proteins on cytosolic calcium concentration distribution fractionally in nerve cells. Interdiscip Sci
3. Zaichick SV, McGrath KM, Caraveo G (2017) The role of Ca^{2+} signaling in parkinson's disease. Dis Model Mech 10:519–535
4. Jha BK, Adlakha N, Mehta MN (2014) Two-dimensional finite element model to study calcium distribution in astrocytes in presence of excess buffer. Int J Biomath 7:1–11
5. Schmidt H (2012) Three functional facets of calbindin D-28k. Front Mol Neurosci 5:25
6. Jha BK, Adlakha N, Mehta MN (2013) Two-dimensional finite element model to study calcium distribution in astrocytes in presence of VGCC and excess buffer. Int J Model Simulation, Sci Comput 4
7. Jha A, Adlakha N (2014) Analytical solution of two dimensional unsteady state problem of calcium diffusion in a neuron cell. J Med Imaging Heal Informatics 4:547–553
8. Naik PA, Pardasani KR (2015) One dimensional finite element model to study calcium distribution in oocytes in presence of VGCC, RyR and buffers. J Med Imaging Heal Informatics 5:471–476
9. Tewari SG, Pardasani KR (2010) Finite element model to study two dimensional unsteady state cytosolic calcium diffusion in presence of excess buffers. IAENG Int J Appl Mathe 40:40-3-01
10. Panday S, Pardasani KR (2014) Finite element model to study the mechanics of calcium regulation in Oocyte. J Mech Med Biol 14:1450022-1-16 (2014)
11. Kotwani M, Adlakha N, Mehta MN (2014) Finite element model to study the effect of buffers, source amplitude and source geometry on spatio-temporal calcium distribution in fibroblast cell. J Med Imaging Health Inform. 4:840–847

12. Pathak K, Adlakha N (2016) Finite element model to study two dimensional unsteady state calcium distribution in cardiac myocytes. Alexandria J Med 52:261–268
13. Dave DD, Jha BK (2018) Delineation of calcium diffusion in Alzheimeric brain. J Mech Med Biol 18:1850028-1-15
14. Magin RL (2010) Fractional calculus models of complex dynamics in biological tissues. Comput Math Appl 59:1586–1593
15. Agarwal R, Jain S, Agarwal RP (2018) Mathematical modeling and analysis of dynamics of cytosolic calcium ion in astrocytes using fractional calculus. J Fract Calc Appl 9:1–12
16. Smith GD (1996) Analytical steady-state solution to the rapid buffering approximation near an open Ca^{2+} channel. Biophys J 71:3064–3072
17. Podlubny I (1999) Fractional differential equations. Academic Press, New York, vol 198, pp 1–366
18. Borak S, Härdle W, Weron R (2005) Stable distributions. In: Statistical tools for finance and insurance. Diss Paper Springer pp. 21–44

Prediction-Based Lossless Image Compression

Mohamed Uvaze Ahamed Ayoobkhan, Eswaran Chikkannan, Kannan Ramakrishnan and Saravana Balaji Balasubramanian

Abstract In this paper, a lossless image compression technique using prediction errors is proposed. To achieve better compression performance, a novel classifier which makes use of wavelet and Fourier descriptor features is employed. Artificial neural network (ANN) is used as a predictor. An optimum ANN configuration is determined for each class of the images. In the second stage, an entropy encoding is performed on the prediction errors which improve the compression performance further. The prediction process is made lossless by making the predicted values as integers both at the compression and decompression stages. The proposed method is tested using three types of datasets, namely CLEF med 2009, COREL1 k and standard benchmarking images. It is found that the proposed method yields good compression ratio values in all these cases and for standard images, the compression ratio values achieved are higher compared to those obtained by the known algorithms.

1 Introduction

In recent years, the availability of huge volume of image data has made the storage and transmission problems very challenging [1]. Many image compression techniques have been reported to solve these problems. Compression can be lossy or

M. U. A. Ayoobkhan (✉) · S. B. Balasubramanian
Computer Science Department, Cihan University-Erbil, Erbil, Iraq
e-mail: mohamed.sha33@gmail.com

S. B. Balasubramanian
e-mail: saravanabalaji.b@gmail.com

E. Chikkannan · K. Ramakrishnan
Centre for Visual Computing, Faculty of Computing and Informatics,
Multimedia University, Cyberjaya, Malaysia
e-mail: eswaran@mmu.edu.my

K. Ramakrishnan
e-mail: kannan.ramakrishnan@mmu.edu.my

S. B. Balasubramanian
Department of Information Technology, Lebanese French University, Erbil, Iraq

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_161

lossless. Lossy compression is mostly used for applications where loss of data may be acceptable [2, 3]. However, lossless compression is needed for many applications such as telemedicine and satellite communication where any degradation of the image quality is not acceptable.

The traditional lossless compression methods normally make use of three steps, namely transformation, data-to-symbol mapping and encoding [4–6]. In this paper, a new lossless compression scheme involving classification, prediction and entropy encoding is presented. The main goal of the proposed lossless compression scheme is to achieve high compression ratio in order to reduce the storage space and bandwidth requirement for transmission and at the same time maintaining the original quality of the images.

As a preprocessing step, the image database is first grouped into sets of similar images. This step leads to a more accurate prediction process which results in achieving a high compression ratio. For classification of images, wavelet-based contourlet transform (WBCT) [7, 8] and Fourier descriptor (FD) [9, 10] is used to extract image features which are then applied to a fuzzy *c*-means (FCM) classifier [11–13]. After classification, each class or group of images is compressed separately using an artificial neural network (ANN) predictor [14, 15]. The predicted pixel values are subtracted from the original values to obtain the residue or prediction error sequence (E). By using an integer format for the predicted and error values, we ensure that no loss is involved in the prediction process. Since these error values only are stored or transmitted, less number of bits are required compared to the original pixel values, thereby achieving compression. In order to improve the compression ratio further, we apply a lossless entropy encoding scheme, namely Huffman coding to the error sequence. For decompression, the error sequence is combined with the predicted values obtained by using an ANN which is identical to the one employed in the compression process.

The proposed compression scheme is evaluated using three datasets, namely COREL-1 k database [16], CLEF med 2009 medical database [17] and standard images such as LENA, BARBARA, etc. An additional feature of the proposed method is that it ensures data security as only prediction errors are involved in the storage and transmission processes instead of the actual image pixels.

2 Proposed Method

The proposed compression method makes use of classification and prediction techniques. For classification, features obtained using wavelet-based countrolet transformation (WBCT) [7, 8] and Fourier descriptor (FD) [9, 10] is employed. Based on these features, the images are classified first into different groups using a fuzzy *c*-means (FCM) classifier [11–13]. The prediction error sequence (E) which is the difference between the original and the predicted pixel values is obtained for each group of images separately [14, 15]. The preprocessing step, namely classification improves the compression ratio as it leads to better prediction thereby reducing the

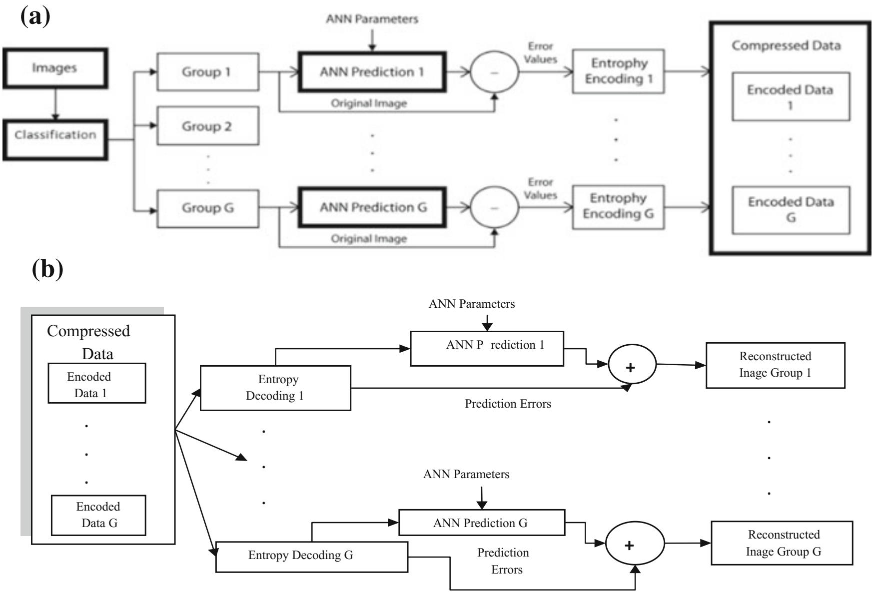


Fig. 1 Block diagram **a** compression stage **b** decompression stage

magnitude of prediction error values. To increase the compression ratio further, the error values are subsequently entropy encoded using Huffman coding technique. For decompression, the reverse steps are adopted. First entropy decoding is done to get back the error values. These error values are then added to the predicted values obtained using an identical predictor (as in the compression stage) to reconstruct the original images.

In the proposed method, the prediction is implemented using artificial neural networks (ANNs). It is assumed that identical ANNs are used for compression and decompression processes. A detailed explanation about the classification phase and the selection of optimum neural network configuration has already been reported [18]. The block diagram of the proposed system is shown in Fig. 1. In the compression scheme (Fig. 1a), the images are first classified into different groups say 1, 2…, G. An optimum ANN network is employed for each group. The error values are then entropy encoded using Huffman coding technique before storing in the memory. The reverse steps are used for decompression of images as shown in Fig. 1b. The error values are combined with the ANN predicted values to obtain the original or reconstructed images. It is assumed that the ANN parameters and the codebooks are available at both the compression and decompression stages.

The ANN parameters shown in Fig. 1 represents the training parameters of each ANN such as the number of inputs, weight values, number of hidden layer neurons, activation functions, etc.

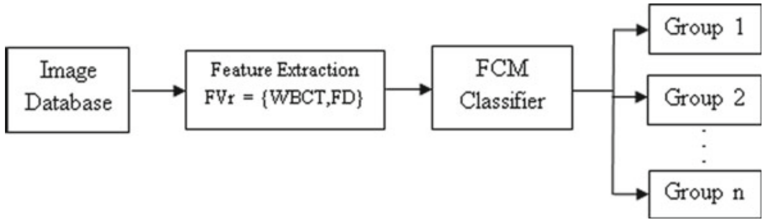


Fig. 2 FCM classification

2.1 *Fuzzy* c*-Means (FCM) Classifier*

As a preprocessing step, classification of images is carried out in order to increase the compression ratio since classification improves prediction accuracy. A fuzzy *c*-means (FCM) [11–13, 18] classifier is employed for this purpose. The feature vector FV_r used for FCM comprises the features obtained using the wavelet-based contourlet transform (WBCT) [7, 8] and the Fourier descriptor (FD) [9, 10]. Figure 2 shows the process of classification in the proposed method.

FCM classification is obtained by minimizing the objective function shown in Eq. (1) [12].

$$A = \sum_{j=1}^{J} \sum_{k=1}^{K} \mu_{ik}^{m} |p_i - v_k|^2 \tag{1}$$

where A is the objective function, J is the number of elements in the feature vector FV_r extracted from the image I, K denotes the number of clusters, μ represents the degree of membership, m is any real value greater than 1, p_i is the ith element in FV_r and $|p_i - v_k|$ is the Euclidean distance between p_i and v_k (centroid of the kth cluster).

2.2 *Feature Extraction*

2.2.1 Wavelet-Based Contourlet Transform (WBCT)

Wavelet-based contourlet transform (WBCT) [7, 8 and 18] is similar to the contourlet transform [19]. WBCT comprises two stages, namely sub-band and angular decompositions. In the sub-band decomposition, separable wavelet filters are used to separate the frequency bands. These filters extract three high-pass bands corresponding to the HH, LH and HL bands. Since wavelet filters are not perfect in splitting high-frequency from the low-frequency components, a fully decomposed directional filter bank (DFB) is used in each band at the second stage. DFB provides angular decomposition using non-separable filter banks [8]. DFB is applied in each high-pass

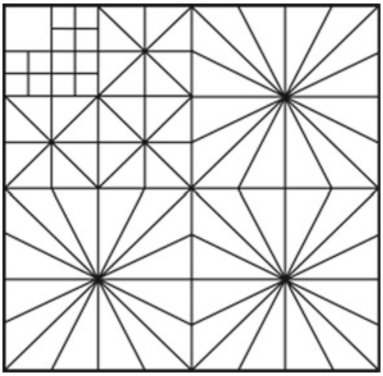


Fig. 3 WBCT with three wavelet levels and eight directions

band at a given level. Starting from the maximum number of directions $D = 2^i$ on the finest level i, where i represents the maximum level of wavelet decomposition. The number of directions depends on the dyadic scale as shown in Fig. 3.

The wavelet coefficients corresponding to the three high-pass bands HH, LH and HL are obtained and used as texture discrimination features. In the proposed method, the coefficients corresponding to LH and HH bands only are employed as texture features. For an image of size $M \times N$, the length of WBCT feature vector WT_n is $N/2$ [8]. In order to improve the classification accuracy, the texture features WT_n obtained using WBCT are combined with the Fourier descriptors (FD_n) which are determined using Fourier transform [5].

2.2.2 Fourier Descriptors

Fourier descriptors (FD_n) are derived by applying Fourier transform on shape signatures [9]. Following are the steps involved in computing the FD_n for an image [10]:

- Boundary coordinates are determined by detecting the boundary points (L).
- Centroid distance function $r(t)$ is obtained from the boundary points.
- Fourier transformation is applied on $r(t)$ to obtain the Fourier coefficients.
- FD_n is determined by using the transformed and original Fourier coefficients.
- The FD_n feature vector of an image with L boundary points comprises only $L/2$ elements because of the symmetry property.

The feature vector FV_r which is applied to FCM classifier is obtained by concatenating WT_n and FD_n. The length of the feature vectors WT_n and FD_n will depend on the size of the image and on the number of boundary points of the image, respectively [5]. For an image of size $M \times N$ with L boundary points, the FV_r is formulated as $FV_r = \{WT_1, WT_2,\ldots, WT_{N/2}, FD_1, FD_2,\ldots, FD_{L/2}\}$.

2.3 Prediction Process

Figure 4 illustrates the steps followed in the prediction process. Figure 4a shows an $M \times N$ image matrix denoted as I where I_{11} to I_{1N} represent the first row and I_{11} to I_{M1} represent the first column. The pixel I_{22} is predicted using the neighbouring pixels I_{11}, I_{12} and I_{21}. Let us denote the predicted pixel as I'_{22}. In the next step, I_{23} is predicted using I_{12}, I_{13} and I'_{22}. This process is continued for the remaining pixels. In effect, the entire image is predicted using the first row and first column pixels as the initial values. Figure 4b shows the predicted image matrix I' corresponding to I. The error matrix of the image denoted as E is obtained as shown in Eq. (2) [18, 20].

$$E = I - I' \tag{2}$$

In I', we make the first row and first column values as zeros so that the first row and first column values of E contain the original pixel values. Figure 4c shows the error matrix E corresponding to I. In order to implement a lossless compression scheme, the predicted values are rounded off to the nearest integers so that the errors (difference between the original and predicted pixel values) will also be integers. This rounding off to the nearest integer operation is carryout both at the compression and decompression stages. This process will ensure that the reconstructed and original images are identical.

2.4 Artificial Neural Network as Predictor

An artificial neural network (ANN) comprising an input layer, hidden layer and an output layer [14, 18] is used as the predictor in the proposed method. The optimum ANN configuration is identified to achieve good prediction accuracy or to reduce the difference between the original and the predicted pixel values. Figure 5 shows the ANN structure with three layers.

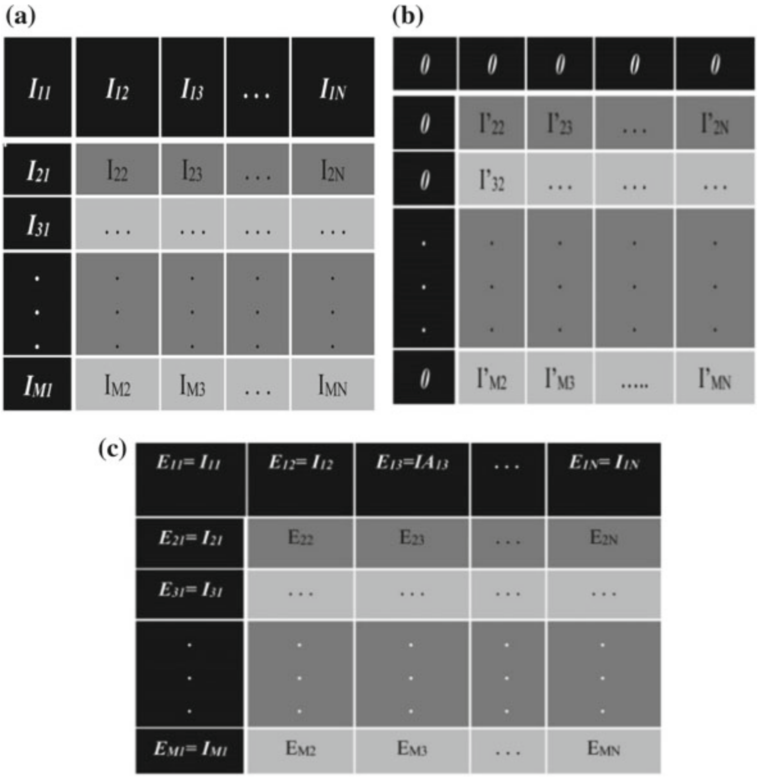


Fig. 4 Prediction steps: **a** original image matrix (I) **b** predicted image matrix (I') **c** error matrix (E)

Selecting an optimum ANN configuration for each group of images plays an important role in improving the prediction accuracy. This process involves finding the optimum number of input and hidden layer neurons, hidden and output layer activation functions and training algorithms. Based on the experimental results, it is found that the optimum number of input and hidden layer neurons is 3 and 10, respectively. The optimum training algorithms, as well as the activation functions, are identified for each group on trial and error basis. The ANN is trained for each group of images separately and for this purpose, 60% of the images in each group are used for training and the remaining images are used for testing [18].

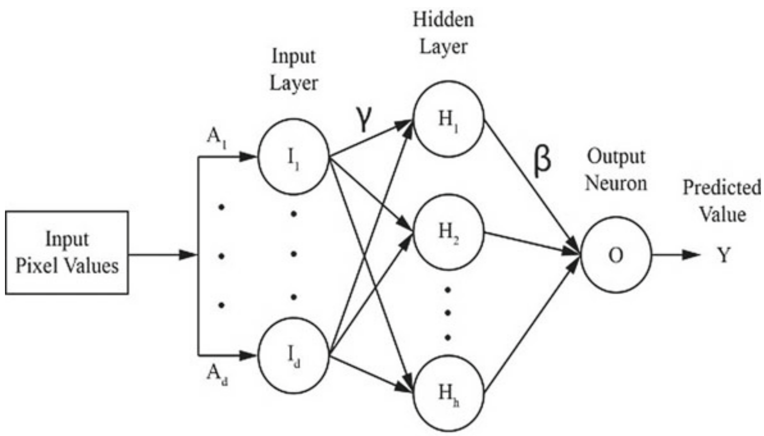


Fig. 5 ANN structure

2.5 Entropy Encoding

In order to improve the compression ratio, entropy encoding is used to compress the prediction errors further. This process, being lossless, the quality of the reconstructed image is not affected. The codebooks used in the encoding and decoding scheme are assumed to be available both at the compression and decompression stages. From our experimental results, it is found that Huffman coding yields better results compared to arithmetic coding. Hence in the proposed method, Huffman encoding is adopted [21].

3 Image Datasets Used for Testing

The proposed system is tested using the following three different sets of image databases.

3.1 CLEF Med 2009 Database

The medical dataset, namely CLEF med 2009 [17] contains more than 15 K images, in which manually 1000 images are selected with 12 different group of images. Figure 6 shows a random sample images from the chest and pelvic girdle group of images from the medical database.

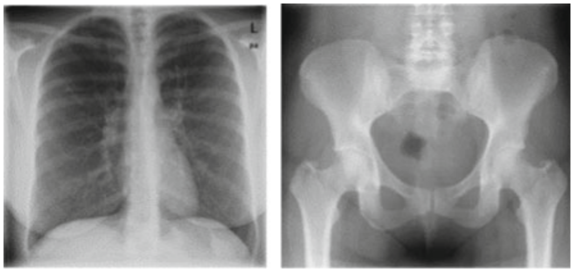


Fig. 6 Sample chest and pelvic girdle images of the CLEF med 2009 database

Fig. 7 Sample images of the COREL database

Fig. 8 Standard benchmarking images

3.2 Corel Database

The general purpose dataset, namely COREL dataset [16] contains 1000 images of 10 groups with 100 images in each group. Figure 7 shows some sample images in the database.

3.3 Standard Images

Figure 8 shows some standard benchmarking images such as Lena, Barbara, Cameraman, Peppers, Man, Goldhill and Boats that are used for testing the proposed system.

Table 1 Compression result of CLEF med 2009 database

Image class	Compression ratio
Chest	5.06
Leg	4.85
Radius and ulna bones	4.95
Pelvic girdle	5.13
Skull front view	5.27
Skull right view	5.76
Skull left view	5.29
Mammogram left	5.95
Mammogram right	6.02
Hand X-ray images	4.85
Pelvic + Back bone	4.87
Neck	4.92
Average	5.24

Table 2 Compression result of Corel 1 K database

Image class	Compression ratio
African people	4.83
Beach	4.95
Building	4.75
Bus	4.65
Dinosaur	5.22
Elephants	4.85
Flower	4.90
Horse	4.73
Mountain	4.69
Food	4.77
Average	4.83

4 Experimental Results

The compression efficiency is calculated using the parameter compression ratio (CR) which is defined as in Eq. (3). The results obtained for the medical and Corel database are shown in Tables 1 and 2, respectively.

$$\text{CR} = \frac{\text{Size of original image (Bytes)}}{\text{Size of compressed image (Bytes)}} \tag{3}$$

The CR value given for each group in Tables 1 and 2 represent the average CR value obtained for the images of that group. Table 3 shows the CR values obtained for standard images using the proposed method. The CR value given for each group

Table 3 Compression result of benchmarking images

STD image name	Image dimension	Compression ratio
Lena	128*128	2.89
	256*256	3.95
	512*512	5.10
Barbara	128*128	3.90
	256*256	5.02
	512*512	6.78
Cameraman	128*128	2.98
	256*256	3.96
	512*512	4.76
Peppers	128*128	2.95
	256*256	4.02
	512*512	5.13
Man	128*128	2.93
	256*256	3.65
	512*512	4.80

Table 4 Comparison of existing methods with the proposed method

STD image name	CALIC	RPC	ETC	Proposed method
Lena	4.48	5.05	4.09	5.10
Barbara	–	6.12	4.58	6.78
Goldhill	4.39	5.10	4.60	5.14
Peppers	4.42	5.05	4.39	5.13
Man	–	–	4.34	4.80
Boats	3.83	4.91	4.12	4.98

in Tables 1 and 2 represent the average CR value obtained for the images of that group. Table 3 shows the CR values obtained for standard images using the proposed method. Table 4 shows a comparison of the proposed method with four existing methods, namely, CALIC [22], ETC [23] and RPC [24]. From Table 4, it is clear that the proposed method yields better CR values compared to other known methods.

5 Conclusions

In this paper, a lossless image compression scheme based on prediction error concept has been presented. Artificial neural network has been employed as the predictor. The prediction process is made lossless by making the predicted values integers. As a preprocessing step, a fuzzy *c*-means classifier using wavelet and Fourier descriptor features has been used. To improve the compression efficiency further, Huffman encoding is performed on the prediction errors. The experimental results show that the compression ratio values obtained by the proposed method are higher compared to those obtained by the known algorithms. Though the method presented in this paper focuses only on grey scale images, the proposed concept can be easily extended to RGB images also.

Ethical Approval For this retrospective type of study formal consent is not required.

References

1. Celik M, Tekalp AM, Sharma G (2003) Level-embedded lossless image compression. Acoustics, speech, and signal processing, 2003. In: 2003 IEEE international conference on Proceedings (ICASSP'03), vol 3, pp III–245, IEEE
2. Ahamed A, Eswaran C, Kannan R (2018) Lossy image compression based on vector quantization using artificial bee colony and genetic algorithms. Adv Sci Lett 24(2):1134–1137
3. Ayoobkhan MUA, Chikkannan E, Ramakrishnan K (2017) Lossy image compression based on prediction error and vector quantisation. EURASIP J Image Video Proc 2017 1:35
4. Bovik AC (2009) The essential guide to image processing. Academic Press
5. Nasir DM, Sayood K (1995) Lossless image compression: a comparative study. In: IS&T/SPIE's symposium on electronic imaging: science and technology. In: International society for optics and photonics, pp 8–20
6. Ramakrishnan K, Eswaran C (2007) Lossless compression schemes for ECG signals using neural network predictors. EURASIP J Appl Sig Proc 2007, 1:102–102
7. Hung CH, Hang HM (2012) A reduced-complexity image coding scheme using decision directed wavelet-based contourlet transform. J Vis Commun Image Represent 23(7):1128–1143
8. Venkateswaran K, Kasthuri N, Alaguraja R (2015) Performance comparison of wavelet and contourlet frame based features for improving classification accuracy in remote sensing images. J Indian Soc Remote Sens 43(4):729–737
9. Zhang D, Lu G (2003) A comparative study of curvature scale space and fourier descriptors for shape-based image retrieval. J Vis Commun Image Represent 14(1):39–57
10. Sokic E, Konjicija S (2016) Phase preserving fourier descriptor for shape-based image retrieval. Sig Process Image Commun 40:82–96
11. Hung CC, Kulkarni S, Kuo BC (2011) A new weighted fuzzy c-means clustering algorithm for remotely sensed image classification. IEEE J Sel Top Sign Proces 5(3):543–553
12. Pandey D, Kumar R (2012) Hybrid algorithm using fuzzy c-means and local binary patterns for image indexing and retrieval. In: Soft computing techniques in vision science. Springer, pp 115–125
13. Fabija´nska A (2009) A fuzzy segmentation method for images of heat-emitting objects. In: Iberoamerican congress on pattern recognition. Springer, pp 217–224
14. Oliveira FD, Haas HL, Gomes JGR, Petraglia A (2013) Cmos imager with focal-plane analog image compression combining dpcm and vq. IEEE Trans Circuits Syst I Regul Pap 60(5):1331–1344
15. Mielikainen J, Huang B (2012) Lossless compression of hyperspectral images using clustered linear prediction with adaptive prediction length. IEEE Geosci Remote Sens Lett 9(6):1118–1121
16. Li J, Wang JZ (2003) Automatic linguistic indexing of pictures by a statistical modelling approach. IEEE Trans Pattern Anal Mach Intell 25(9):1075–1088
17. Tommasi T, Caputo B, Welter P, Guld MO, Deserno TM (2009) Overview of the CLEF 2009 medical image annotation track. In: Workshop of the cross-language evaluation forum for european languages. Springer, pp 85–93

18. Mohamed Ahamed A, Uvaze C, Eswaran R Kannan (2017) CBIR system based on prediction errors. J Inf Sci Eng 33(2):347–365
19. Li H, Chai Y, Li Z (2013) Multi-focus image fusion based on nonsubsampled contourlet transform and focused regions detection. Optik-Int J Light Electron Opt 124(1):40–51
20. Ayoobkhan MUA, Chikkannan E, Ramakrishnan K (2018) Feed-forward neural network-based predictive image coding for medical image compression. Arabian J Sci Eng 43(8):4239–4247
21. Abo-Zahhad M, Gharieb RR, Ahmed SM, Abd-Ellah MK (2015) Huffman image compression incorporating dpcm and dwt. J Sig Inf Proc 6(02):123
22. Wu X, Memon N (1997) Context-based, adaptive, lossless image coding. IEEE Trans Commun 45(4):437–444
23. Zhou J, Liu X, Au OC, Tang YY (2014) Designing an efficient image encryption-then compression system via prediction error clustering and random permutation. IEEE Trans Inf Forensics Secur 9(1):39–50
24. Liu W, Zeng W, Dong L, Yao Q (2010) Efficient compression of encrypted grayscale images. IEEE Trans Image Process 19(4):1097–1102

Audio and Video Streaming in Telepresence Application Using WebRTC for Healthcare System

Dhvani Kagathara and Nikita Bhatt

Abstract Currently, health care in Lower- and Middle-Income Countries (LMICs) is suffering from shortage of trained physicians in rural areas. The development of certain technologies, particularly in sensors has yielded the birth of mHealth. The main objective of this paper is to develop a telepresence application for a mobile platform. The functionality of the system will include the ability for the doctor to connect with a doctor/patient from a remote location using mobile or web application, and the ability to diagnose patient disease remotely and recommend follow-up care. There is a requirement for low price, well suited, and easy-to-use video communication system. The web real-time communication (WebRTC) permits browsers to set up a peer connection to deliver information and media with real-time conferencing capabilities through simple JavaScript APIs.

1 Introduction

The improvement of financial and social elements regarding a country is said to be complementary to each other. The fields including education or health care are suffering from the social need which sooner or later depicts the financial growth and in the long run, the excellent of lifestyles. India has expected digital adoption, the Indian Healthcare marketplace, that is certainly worth around US$ a 100 billion, will probably develop at a CAGR of 23% to US$ 280 billion through means of 2020 [1]. Including infrastructure or medical specialists via myself will now not be capable about unravel India's considerable unmet desires among health care. It desires to be supported through the era. Scientific technology innovation may remain the tool to redact current care accessible, available then less costly to all by way on lowering the cost of the product or delivery.

D. Kagathara (✉) · N. Bhatt
Chandubhai. S. Patel Institute of Technology, Charotar University of Science and Technology, Nadiad, India
e-mail: kagatharadhvani@gmail.com

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_162

The characteristic of playing back audio and video also as the record is Wight received is recognized as streaming that is a point in accordance with point and multipoint. Unlike everyday data file, a streaming media file for consideration is big, as a consequence calls because of excessive channel bandwidth. Furthermore, streaming media moreover contains severe demand inside the timing of bundle delivery. It has some constraints consisting of bandwidth, latency, or congestion. Video streaming may stay categorized into two categories such as live and on-demand. In this paper, we observe nearly used streaming protocols for mobile utility or web [2]. They may additionally be lower layer as TCP or UDP and upper layer so RTP, RTCP or RTSP [3]. In video streaming, lower layer protocol such as Transmission Control Protocol (TCP) or User Datagram Protocol (UDP) is not feasible to use [3]. So far, we will use the higher layer protocols such as Real-time Transport Protocol, (RTP) Real-time Control Protocol (RTCP), and Real-time Streaming Protocol (RTSP) [3]. Real-Time Communication (WebRTC), a project maintained by way of means over Google among 2011 primarily based on RTP is analyzed or related according to such through IETF and its browses API by the usage of W3C [4]. This gadget allows ongoing correspondences by means of a couple which use Programming Interfaces (API) alongside a top-notch, low idleness, and low transmission capacity utilization [2].

2 WebRTC

WebRTC is the collection of protocols, standards, and APIs [2]. Audio, video and information shares between peers using the connection [2]. There is no software or any extra plug-ins required [5]. The WebRTC implementation is created through Google and open source launched it [2]. World Wide Web Consortium (W3C) organization standardized the WebRTC at the application level and Internet Engineering Task Force (IETF) organization standardized the WebRTC at the protocol level [4].

The major components of WebRTC are:

2.1 Media Stream

Get user media allows to acquiring audio or video from the hardware such as camera, microphone, screen capturing, or display [6]. Using get User media API, we can get the screenshots and share the screen with different parties also [6] (Table 1).

Table 1 Media engine performance [10]

Module	Measured statistics
Video capture	Captured video resolution, and a number of frames captured per second [10]
Video encoding	Encoded stream bit rate, frame rate and resolution also encoder ramp-up speed [10]
Video decoding/rendering	Frame numbers decoded/rendered per second; resolution of the rendered stream [10]
Networking	Bytes/packets sent, received and dropped, RTT [10]
Jitter buffer	Jitter, A/V Sync [10]

2.2 RTC Peer Connection

RTC peer connection is used for setting the audio/video streams [2]. It consists a lot of special tasks like codec execution, signal processing, bandwidth administration, safety streaming, and many more which represents the contain into MediaStream and DataChannel by presenting a handshake mechanism because of pair machines according to exchange fundamental data so a Peer-to-Peer connection may stay set up [6] (Table 2).

2.3 RTC DataChannel

Connected users can share audio, video, and data using RTC DataChannel [2]. Which is used for Bidirectional communication among peers and exchange the data which is in any format [2]. Which is share arbitrary data between peers [6] (Table 3).

2.4 GeoStats

API call that hat permits getting diverse insights about a WebRTC session [2].

3 WebRTC Architecture

Figure 1 illustrates the WebRTC overall structure which API used and different audio and video codecs used in the streaming application. Also, it shows the features supported by browsers and application for real-time communication [7]. WebRTC offers net software builders the potential to put in writing wealthy, real time multimedia

Table 2 Connection flow measurements [10]

	Parameters	Measurement statistics
Initialization	New peer connection	Time for instantiating a new RTC peer connection object [10]
	Get user media	Time spent capturing user media excluding time waiting for user permission [10]
	Can Play Local Media	Time spent after media request until local media becomes available [10]
	Setup signaling channel	Time to setup signaling channel using web sockets [10]
Communication	Make call	Time for setting session descriptions and creating an offer [10]
	ICE hole punching	Time spent on ICE hole punching process [10].
	Make answer	Time for setting session descriptions and creating an answer [10]
	Signaling	Time spent propagation of answer/offer messages and signals across the network [10]
	Open dataChannel	Time for the opening dataChannel [10]
	Play remote rtream	Time for play remote media stream to be available [10]

Table 3 DataChannel measurements [10]

Category	Measured statistics
Network	Bytes sent/received/RTT [10]
Application	Messages sent/received/RTT [10]

applications (suppose video chat) at the web without requiring plug-ins, downloads or installs. Its cause is to help construct a robust RTC platform that works throughout more than one web browsers throughout a couple of platforms [8].

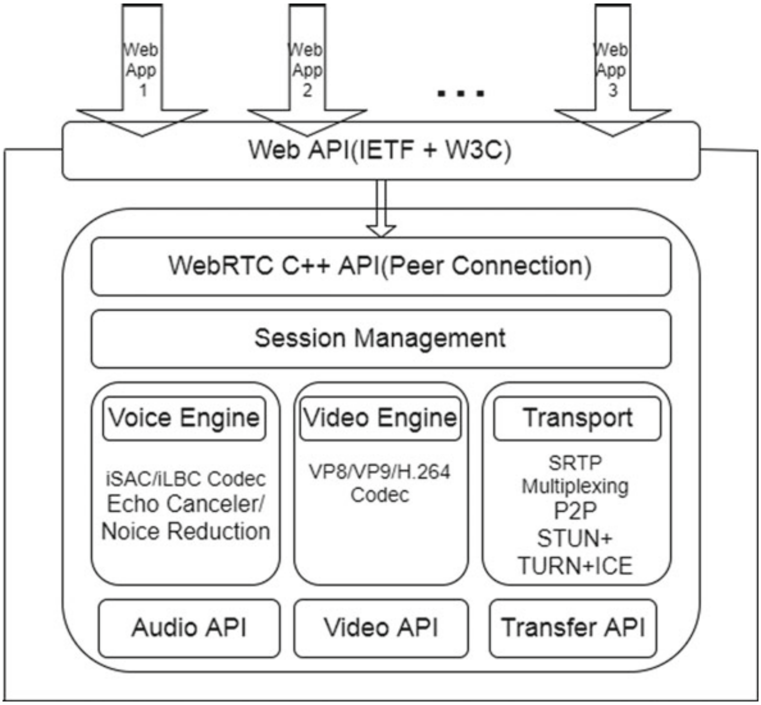


Fig. 1 WebRTC structure

3.1 Voice Engine

The voice engine is usually called sound engine [7]. They are processed for echo cancelation and noise reduction [9]. The sound machines facilitate the audio codecs which is optimized by narrowband and wideband [9]. The concurred sound codecs up to desire require in congruity with be upheld by method for WebRTC all around coordinated programs comprise of iLBC (i.e., a narrowband discourse codec for VoIP or gushing sound), iSAC (i.e., a wideband voice codec in light of the fact that VoIP at that point spilling), or Opus (i.e., a codec so much aides bit rate encoding) [7]. In sender side, they process the raw streams to enhance the quality and match the bandwidth and latency which is continuously fluctuating [9]. From receiver side, received streams decode and adjust the latency delay and network jitter [9].

3.2 Video Engine

The video engine is a framework as offers together with video streams that come beside the digital camera in conformity with the community and from the community in accordance with the show display [7]. Just as the frame engine, the video engine helps the video codecs. VP8 is writing in time agreed with the video codec [7]. It helps potential jitter buffer because of the video who allows in accordance with hide the outcomes concerning get rid of jitter then packet loss in imitation of enhance the overall video top-notch [7]. It additionally allows image enhancement, as much an instance, eliminating video noise from the image capture by using the usage of behavior regarding the webcam [7]. It is designed for low latency so that it is well suited for RTC [9].

3.3 Transport Support

At the transport layer pass, the audio and video streams to the peers using this activity [7]. TCP buffers all packets after th intermediate packet is lost and wait for retransmission after delivering the stream [9]. A secure Real-Time Protocol is used with WebRTC to secure the streams of audio or video using encrypts of all the records [7]. UDP delivers each packet at the moment it arrives which is no promises on ordered data [9].

3.4 Session Management

The session management defines as beneficial to WebRTC as abstract signaling approach and the protocol according to session data as explained or leaves the statistics about the signaling implementation on the hand concerning application builders [7]. Session description protocols are most affected by WebRTC [7]. C++ native API of WebRTC at this API layer allows browser markers in accordance with except issues put into effect the web API because of his or her browses [7]. Firewalls deals the parameters for every stream and implement congestion control, flow control and encryption of data, and more [9].

4 Implementation

4.1 Environment Setup [7]

We are using the WebRTC functionalities for establishing the environment which is established previously by EasyRTC APIs and JavaScript APIs for Web Browsers. EasyRTC is the open-source project [7]. We used the node.js server and JavaScript APIs to run the application in web browser and for the mobile application, we used the WebView [7].

4.2 Audio and Video Streams Gathering [7]

For audio and video streams gathering found complexity [7]. Call EasyRTC. getLocalStreame and EasyRTC [7]. set video ObjectSrc [7]. We have to invoke the method for ask the permission for Local camera and microphone [7]. After giving the camera and microphone permission, the user can see their local screen. For incoming streams, we have to call the callback method [7]. For setting the bandwidth for each video track we can use EasyRTC setVideoBandwidth() [7].

4.3 Signaling and Peer Connection [7]

To establish the communication, first exchange the data between peers [7]. As an example, for making the connection in the same network, only public IP address is necessary [7]. But the process is complicated when peers were far to each other and peer's public IP addresses and port address are hidden to each other [7]. If one of the peers is not available, we have to first acquire the viable IP addresses and port candidates for each peer, traverse the NATs, and then run the connectivity [7]. The single interface encapsulating all the connection setup, control, and state PeerConnection API within the WebRTC is dealing with the life cycle for every peer connection [7].

Figure 2 explains the signaling interactions and the requirements between two peers.

(a) Whenever the user from one side which uses browser 1 clicks the "join" button, the browser of user obtains and presentation the local media [7].
(b) After a user of one side wants to connect to the user of another side which uses browser 2 for video conference of the secure channel using the signal server [7].
(c) After the session is mounted between both users [7].
(d) For exchange, the media, codecs and setting their bandwidth information, a user from one side uses the Session Description Protocol (SDP) [7]. Real media itself is not connected via offer [7].
(e) As soon as the offer is gotten, a client from opposite side makes the offer to associate with the client in one side and furthermore sends the comparing session profile utilizing the SDP [7]. After that, the client from one side realizes that client from the opposite side is prepared to associate utilizing shared connection [7].
(f) User from the opposite side makes an intuitive network foundation (ICE) operator, (ICE A) to the client from program 1 side [7]. The ICE specialist finds the closer IP with the port tuple of companion and uses outer STUN server for get general society IP and port tuple [7].
(g) The client from program 2 side accepting the ICE an indistinguishable activity from a client from program 1 side to make and send ICE specialist (ICE B) [7].

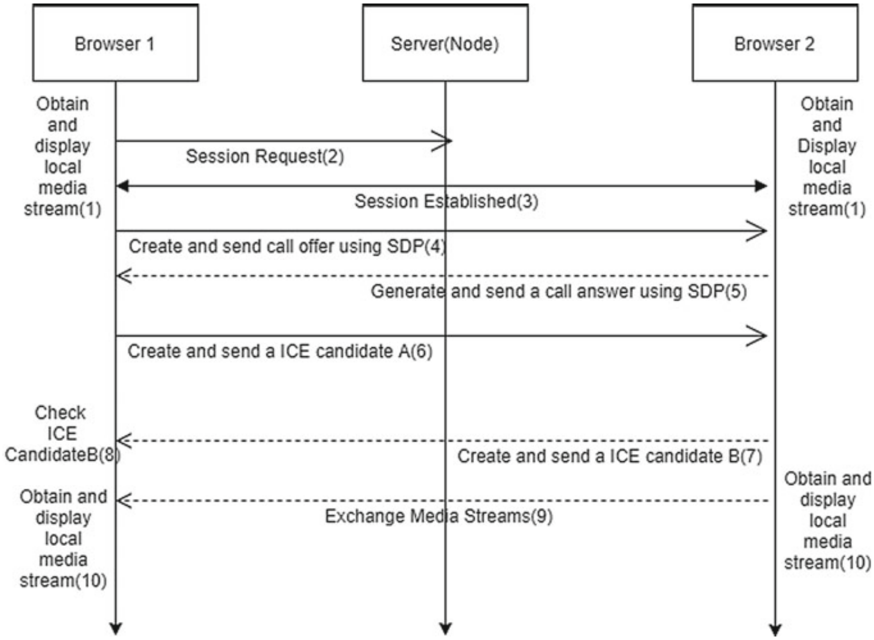


Fig. 2 Peer-to-peer interaction

(h) TURN server utilized as an intermediary to hand off movement if clients cannot make peer association utilizing STUN server [7]. It will get data ahead of time which is suits in ICE B [7].
(i) User from one side sends the video stream to another side user [7]. Both side data and nearby media streams established [7]. In the wake of getting general society IP locations and port assortment tuples ahead of time from ICE A [7].
(j) In the end, media content is shown by both the browsers and the video conferencing is operating [7].

STUN/TURN servers can be used to handle a small number of participants. Javascript code is added to make the peer connection [7].

4.4 Chatting Room Establishment [7]

This is used for the patients and doctors to conduct the meetings [7]. Codecs for client and server needs to be implemented. Using Socket.io, the first client is getting the chat channel [7]. As the connection is established, the client can send the text. After every user joins the room. Display the incoming messages and outgoing messages [7].

5 Limitations

Limitations of the existing telepresence applications are: Present video streaming application used for telepresence criticized for plenty of reasons: (a) they are often too high priced to buy and preserve, (b) using technologies which is proprietary cannot be compatible with each other, and (c) to keep the device they require fairly skilled IT personnel [7].

6 Conclusion and Future Extension

In this paper, we have discussed the WebRTC standard-based application for telehealth which is simple, interoperable, and inexpensive for video conferencing. The proposed healthcare system will be helpful for communication of doctors and patients. The system would be developed using web real-time communication with better quality audio and video streams using different protocols and standards to evaluate the need for Telehealth. Real-time communication use for all the platforms of a web on cloud computing and Android/iOS devices. We have discussed related implementation of a WebRTC based system which is emerging different protocols and architecture. Which is provide the good user experience in web and mobile both. We have used different mechanisms like RTP based media exchange and peer-to-peer data exchange. WebRTC provides more developer friendly environment that does not require specific software and plug-ins, and it will be available anytime and anywhere. WebRTC requires permission to use camera and microphone to access from the user. WebRTC reduces energy intake and bandwidth utilization and accordingly enhances the effectiveness of the system. We have used different approaches for real-time communication but the methods have some limitations which may be overcome by improving an existing system. In the future, healthcare device can handle multiple connections with low bandwidth and high latency. Audio and video must of good quality.

References

1. https://www.ibef.org/industry/healthcare-india.aspx
2. Ivan santos-gonzalez A (2016) R.-G.-B.-G.-G: real-time streaming: a comparative study between RTSP and WebRTC. Springer, pp 313–325
3. Arsan T (2013) Review of bandwidth estimation tools and application to bandwidth adaptive video streaming
4. https://en.wikipedia.org/wiki/WebRTC
5. Jansen BA (2016) Performance analysis of WebRTC-based video conferencing. Springer
6. Laurent Lucas HD (2017) USE together, a WebRTC-based solution for multi-user presence desktop. Springer
7. Julian Jang-Jaccard SN (2014) WebRTC-based video conferencing service for telehealth. Springer, CSIRO Computational Informatics (CCI), Marsfield, Australia

8. https://webrtc.org/architecture/
9. https://princiya777.wordpress.com/2017/08/19/webrtc-architecture-protocols/
10. Sajjad Taheri LA (2015) WebRTCBench: a benchmark for performance assessment of WebRTC implementations. IEEE

Review on Image Segmentation Techniques Incorporated with Machine Learning in the Scrutinization of Leukemic Microscopic Stained Blood Smear Images

Duraiswamy Umamaheswari and Shanmugam Geetha

Abstract This paper is a contemplated work of N-different methods that have been employed in the area of revealing and classifying leukocytes and leukoblast cells. Blood cell images obtained through digital microscopes are taken as input to the algorithms reviewed. In bringing out the nucleus and cytoplasm of White Blood Cells (WBCs), the images have been undergone by a variety of image segmentation techniques along with filtering, enhancement, edge detection, feature extraction, classification, and image recognition steps. Apart from image processing, the analysis and categorization of the leukemic images are handled using some other machine learning techniques of computer science discipline. Assessment of accuracy and correctness of the proposals were done by applying texture, color, contour, morphological, geometrical, and statistical features.

1 Introduction

Biomedical image processing is a prominent field of computer science for processing medical images which are digital in character. Healthcare industries are facing hard issues in detecting the cancerous malignant cells in early stages. In that way, leukemia is a grievous disease that threatens human by reducing their lives.

Leukemia damages the leukocytes of blood without exhibiting the causes in prior. White blood cells otherwise called as leukocytes, the antibodies of the human immune system, get spoiled by leukemia, which in turn causes continuous illness. Numerous research processes have been performed to detect and classify malignant leukocytes

D. Umamaheswari (✉)
Department of Computer Science, Vidyasagar College
of Arts and Science, Udumalpet, Tamil Nadu, India
e-mail: ums1082@gmail.com

S. Geetha
Department of Computer Science, Government Arts College,
Udumalpet, Tamil Nadu, India
e-mail: geet_shan@yahoo.com

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_163

and regular WBC. Cytoplasm and nuclei of the WBC are needed to be separated, for processing in the recognition of leukemia cells. For this cause, different sorts of computer-aided techniques are proposed. Eventually, it is obvious for the researchers getting assistance from pathologists and oncologists for the error-free classification. Microscopic images of blood smear are being the major source of input and because there are no standard datasets available, it is must perform sensitive evaluations and comparisons of one's method with others in terms of accuracy.

In this paper, a review work is made about various techniques that have been incorporated in conjunction with Digital Image Processing (DIP) methodologies in the arena of leukemia detection. The paper describes leukemia and its types in Sect. 2, various stages of DIP are discussed in Sect. 3, review of the research activities made in the discipline of leukemia diagnosis is done in Sect. 4, Sect. 5 explains the issues faced by the researchers, and the manifestation of the review work is concluded in Sect. 6.

2 Leukemia and Its Subtypes

Leukemia is also acknowledged as the blood cancer that affects the spongy substance surrounded with the bone marrow, which is in charge of producing blood cells. Leukemic bone marrow secretes surplus amount of blood cells rather what is needed. Leukemia is recognized with diverse types. With the conduct of the proper examination of leukocytes, the nature of leukemia can be discriminated. Research activities of this line will be useful to the hematologists in reporting the results of blood tests. Major categorization of leukemia is of four types depending on whether it is acute otherwise chronic, leukoblastic, or else myeloid: Acute Lymphoblastic (Lymphocytic) Leukemia (ALL), Acute Myeloblastic (Myeloid) Leukemia (AML), Chronic Lymphoblastic Leukemia (CLL), and Chronic Myeloblastic (Myeloid) Leukemia (CML).

Most of the diseases are diagnosed by examining the blood samples through the microscopes. Hematologists sense the illness with the application of multifarious features of blood; one among them is counting of blood cells. In the medicinal field, numerous research activities are being in progression to bring out the exact count of blood cells. The differential count is the common label used to represent the count of WBC. Blood is the combination of various components: plasma, erythrocytes, leukocytes, and platelets. The last three are contained in plasma, a cell-less watery substance. Erythrocytes are widely referred to the character name red blood cells. The immune system of human being provides resistance from diseases by means of the antibodies referred to as leukocytes or WBC. WBCs are categorized about more than 20 types. Five significant subtypes are, namely, neutrophil, basophil, eosinophil, monocyte, and lymphocyte as shown in Fig. 1.

Technology developments in the counting of WBC and their subtypes also help in the healthcare industry in diagnosing the health disorders. Researchers support the biomedical technicians to perceive the leukemia in the way of proposing accurate

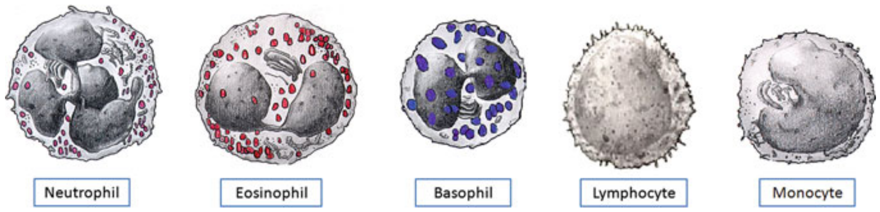


Fig. 1 Subtypes of WBC

algorithms intended for partitioning the nuclei or cytoplasm of the blood cells as well as classifying them as normal or malignant cells.

3 Stages of Image Processing

Image processing is about processing digital images in a series of stages. According to our current study of leukemia detection and classification, the stages of DIP shown in Fig. 2 are carried out in the subsequent manner:

1. Image Acquisition: It is the task of getting microscopic blood images or biopsy microscopic images.
2. Image Transformation: The images are converted to a standard format and color space.
3. Image Restoration: Microscopic images with varying noise types are made better, without taking into account the intended features of the image.
4. Image Filtering: The biological or acquisitional artifacts are cleaned up with the aid of filters.
5. Image Enhancement: Images can be viewed better by enhancing their color, contrast, and brightness like features.
6. Edge Detection: Edges are the boundaries of the components which are found with edge detectors.
7. Morphological Operations: To extract the shape and size features of cells of blood, opening, closing, dilation, and erosion like operations are carried out.
8. Segmentation: This is the process of partitioning nucleus and cytoplasm of WBCs.
9. Feature Extraction: Features are the characters of an image. After segmentation of the sub-images, the texture, color, contours, geometrical, and statistical features are extracted from them.
10. Object Recognition: This process results in naming the parts of microscopic images.

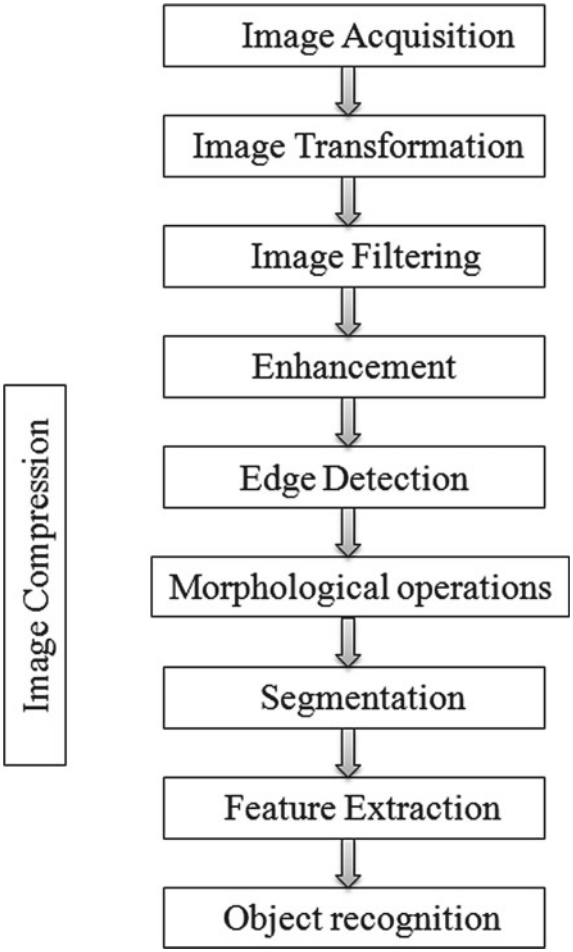


Fig. 2 Stages of digital image processing

4 Review of Literature

The baseline of the research activities on identifying leukemia is partitioning the nucleus and cytoplasm of the WBC. This work is implemented with the support of number of image segmentation algorithms. The segmented regions are taken into contemplation for categorizing different leukemia types with the help of machine learning techniques like classification.

4.1 *Image Segmentation*

Segmentation does the vital task of separating the nucleus and cytoplasm for analysis. The algorithms come up with several flavors for effective segmentation.

4.1.1 Segmentation by Thresholding

Cseke [1] projected a method of segmentation for the automatic differential count of WBC with the application of Otsu's threshold and automatic thresholding with the segmentation probability of 0.92. Liao and Deng [2] have presented a segmentation scheme seeking the count of WBC by applying thresholding with shape detection method in a step-by-step manner. Prasad et al. [3] presented an algorithm which performs recursive segmentation by comparing an arbitrary threshold value with the histogram of pixel intensities, which stripes off background pixels by segmenting only relevant pixels. Then, single cells are classified from cell clusters using signature plots of cluster classification; lastly, cell size, count, and location are determined and showed an accuracy of 97%. The authors Prasad with Badawy [4] proposed an algorithm for identifying, classifying, and quantifying the count of leukemia cells. Here, the pixels under ROI are recursively segmented by comparing the intensity values with a threshold and gives 95% accuracy for poor quality images yet.

Biswas and Ghoshal [5] made an increase in the performance of Sobel edge detector by a succeeding thresholding assessment-based watershed algorithm. The evaluation of this algorithm is done on 30 microscopic images and returns the accuracy of 93%. Kit et al. [6] proposed a framework for the Complete Blood Count (CBC). In this work, red and WBCs are extracted by the way of color space analysis, thresholding, and Otsu's method. Afterward, the appropriate cells are counted using topological structure analysis; then, cell mass regions are predicted through Hough Circle Transform (HCT), and the accuracy level is 92.93% in RBC count and 100% in WBC count.

Scotti [7] reported methods to measure the attributes of WBC. In their proposal, segmentation of WBC is done by gray-level thresholding. The combination of Gram-Schmidt technique and parametric deformable representation was introduced by Rezatofighi et al. [8] for segmenting the WBC's nuclei and cytoplasm, and also a proper thresholding technique was applied to improve the outcome of segmenting cytoplasm. In [9], Bergen et al. reported a scheme to segment the erythrocytes (RBCs) and leukocytes (WBCs). Otsu's thresholding is applied to separate nucleus. Rezatofighi et al. [10] proposal is based on orthogonality principle added with the Gram-Schmidt method which amplifies the desired color vector and suppresses the undesired color vectors for segmenting the nucleus of WBC with suitable threshold. Sadeghian et al. have designed a framework [11] with the intention of segmenting nucleus and cytoplasm. Nucleus segmentation was carried out through GVF snake, and cytoplasm was found through Zack thresholding with the segmentation accuracy of 92 and 78%, respectively.

In the paper [12], Nor Hazlyna et al. mentioned a segmentation methodology to segment blood image of leukemia into blast region and background region by means of the color spaces HSI and RGB. This is implemented with a threshold value. Mohamed and Far [13] proposed a new automatic segmentation algorithm that used Otsu's threshold to convert the contrast-enhanced gray image into twofold binary image. Its segmentation accuracy is 80.6%. Di Ruberto and Putzu [14] have presented an efficient segmentation method based on fuzzy set thresholding. Ahasan et al. proposed an algorithm [15] for segmenting the nuclei of blood white cells; it has a number of stages such as color space conversion, color thresholding, binary imaging by Otsu, average filter, marker-controlled watershed, different morphological operations, and eliminating borders with an accuracy of 88.57%. Lina et al. [16] presented a scheme to pin out WBCs. In this work, color filtering is prepared by comparing with a threshold value. The method works with the leukocyte detection accuracy rate of 82.12%. In [17], the authors Ananthi and Balasubramaniam implemented the segmentation by choosing a threshold from a fuzzy set with maximum similarity.

Sandhu [18] made a study on leukemia detection and concluded that watershed segmentation with optimal thresholding is best with the segmentation accuracy of 99.85, 99.2, and 99.63% for segmenting cells, nucleus, and cytoplasm, correspondingly. Putzu together with, Di Ruberto [19] introduced a methodology to find out WBCs and their groups by using triangle-based thresholding method of segmentation. Further, the cytoplasms along with nuclei of WBC are found. Gosh et al. demonstrated [20] a scheme of revealing leukemia with fuzzy deviation and adapted thresholding methods. Khobragade et al. [21] used Otsu's thresholding to convert the gray image into binary. From this, it could be possible to classify leukemia types with 91% accuracy. Abbas and Mohamad [22] proposed a new method of segmentation. In this method, the necessary areas are found by binarizing the image using Otsu's threshold with the segmentation accuracy of 96.56% with a reduction in computation time about 50%. Khashman and Al-Zgoul [23] proposed a scheme of segmenting the cytoplasm and nucleus regions with the assistance of bimodal thresholding. It assures the segmentation accuracy ratio of 98.33%. Abbas et al. [24] introduced a new scheme of segmentation assisted with a threshold value. The method improved the accuracy to 0.8955%. Li et al. [25] presented a technique of WBCs segmentation in ALL images with the application of the dual-threshold method. This proposal attained 97.85% of segmentation accurateness.

4.1.2 Segmentation by Watershed Algorithm

Nilsson and Heyden [26] introduced a model of segmentation method using watershed algorithm and region growing for segmenting the nucleus and accordingly its cytoplasm. This work gives the differential counts of WBC. Jiang et al. [27] offered a scheme to segment WBCs. In this scheme, watershed clustering is applied in separating cytoplasm. Shankar et al. [28] explained a method to identify ALL by a set of image processing steps including segmentation with watershed algorithm; afterward, abnormal cells are separated. This method attains more than 90% of accuracy. Ghane et al. [29] found a technique combining thresholding, k-means clustering, and watershed algorithms to segment both the blood cells and nucleus of lymphocytes. The work was compared with manual segmentation by three factors, namely, similarity measures, sensitivity, and precision.

Belekar and Chougule in [30] explored a scheme to segment nuclei and cytoplasm of leukocytes by utilizing watershed transform along with level set. Liu et al. [31] demonstrated a method to segment WBC using average shift clustering along with watershed algorithm.

4.1.3 Segmentation by k-means Clustering

Sinha and Ramakrishnan [32] worked on color images of discolored peripheral blood smears to reveal various classes of WBCs. The system does the segmentation in two stages: Classification by k-means clustering and feature extraction by EM algorithm. Basima and Panicker [33] presented an automated method of detecting, counting, and classifying blood white cells. In this system, WBCs are detected through k-means segmentation; features are acquired by using gray-level co-occurrence matrices, and then classification of healthy WBCs and blast WBCs by SVM with the accuracy range of 94.56%.

Foran et al. [34] introduced a computer-aided medical decision support system that differentiates the hematological malignancies using. In this system, segmentation is done by clustering. Su et al. [35] proposed a technique to count blast, erythroid, lymphocyte, monocyte, myeloid, and neutrophil cells in separate in addition to the overall cell count. Using these details, AML is detected by segmenting the cells using k-means clustering. Laosai and Chamnongthai [36] developed a system to segment and classify lymphoid and myeloid blood cell. Segmentation of nucleus is done by k-means clustering and SVM is used for classification. The accuracy offered by this system is 92%. In [37], a framework was proposed by Kumar et al. with the purpose of identifying and classifying cancer using biopsy microscopic images. It involves a number of steps including k-means segmentation.

Neoh et al. [38] introduced a decision support system to identify ALL by using a clustering algorithm in conjunction with Simulating Discriminant Measures (SDMs) in segmenting the nucleus and cytoplasm of both lymphocytes and lymphoblasts. The system has used classifiers such as Multi-Level Perceptron (MLP), SVM, and Dempster–Shafer. Among these, Dempster–Shafer brought out the accuracy rate of 96.72%. Agaian et al. [39] presented a technique to make known the ALL by segmenting the nucleus with the application of k-means clustering as a segmentation methodology. This application showed the accuracy rate of 98%. Moradi Amin et al. [40] give away a scheme of revealing ALL with the help of k-means clustering for segmentation and SVM for classification. Hazra et al. [41] achieved 90% of accuracy in segmentation with the application of k-means clustering. Indira et al. [42] exercised k-means clustering algorithm for the segmentation of WBCs' nucleus to ascertain AML. SVM classifies the cell with the assistance of features gathered by LBP and

LTP. Madhukar et al. [43] proposed a method to find AML; in that, segmentation is done with k-means clustering and classification by SVM with the accuracy rate of 93.5%. Soltanzadeh et al. [44] used k-means segmentation for separating the nuclei of leukocytes and curvelet transforms to enhance the extracted region. This method provides 80.2 and 84.3% of specificity and sensitivity.

Sajjad et al. [45] and Selvaraj and Kanakaraj [46] partitioned nucleus of WBC using k-means clustering algorithm. Sivakumar and Ramesh [47] used the k-means clustering along with morphological actions to segment the leukocytes. Kazemi et al. [48] proposed a method to reveal AML by applying k-means technique of segmentation of WBCs' nucleus and SVM for classification of subcategories of AML. Sharma and Kinra [49] examined k-means method and histogram equalization to extract and count the healthy leukocyte and malignant leukoblast cells for the early detection of leukemia. Harun et al. [50] combinedly used k-means and fuzzy c-means along with moving k-means clustering algorithms in segmenting the injured blast cells. Among the three methods, MKM produces the best segmentation sensitivity rate of 86.18%.

4.1.4 Segmentation by Morphological Operations

Piuri with Scotti [51] both worked on microscopic color images for automatic detection and categorization of leukocytes. In their work, leukocytes separated from all the other cells of blood using morphological operations, and the subtypes of WBC such as basophile, eosinophil, neutrophil, lymphocyte, and monocyte are classified using neural classifier. Theera-Umpon along with Dhompongsa [52] examined that the nucleus of WBC is with mathematical morphology. Bayes' classifiers and neural networks are functioned as classifiers in fivefold, which achieved 77% of classification rate.

Vogado et al. [53] proposed a leukemia segmentation technique using morphological processes. Results of experiments have figured out the kappa index of 0.9306, 0.8603, and 0.9119 with the databases ALL-IDB 2, BloodSeg, and leukocyte, correspondingly. In [54], Bouzid-Daho et al. introduced a system; in that, the abnormal cells are detected by applying mathematical morphological operators for segmentation. Bhattacharjee [55] with Saini developed a GUI using the morphological operators for partitioning nucleus through segmentation. The classification was implemented through K-nearest neighbor, k-means clustering, Artificial Neural Networks (ANNs), and SVM. In [56], Scotti proposed a system that works on peripheral blood microscopic images. Here, the leucocytes are separated from other blood cells, lymphocyte cells are extracted, and morphological indexes are evaluated from them; finally, the existence of leukemia is classified. Vishwanathan [57] presented a new approach for detecting leukemia by fuzzy C-means cluster optimization based on morphological contour segmentation.

Gumble and Rode [58] have investigated DIP techniques such as median filtering, Sobel edge detector, morphological erosion and dilation for segmentation of nuclei and cytoplasm of WBC, morphological features extraction, and KNN classifier for classifying normal and malignant blood cells. Polyakov and Nikitaev proposed [59] a morphological feature extraction method to differentiate blast and normal lymphocytes. The features are measured using statistical measures like standard deviation and mean. Grimwade et al. reported [60] the application of image flow cytometry with microscopic blood smear images to detect acute leukemia. In this streamline, the components of cell are identified by morphology. Singhal along with Singh [61] distinguishes normal leukocyte cells and leukoblast cells by applying morphological dilation. Then Local Binary Pattern (LBP) features of texture are acquired. It brought out fine accuracy while classification.

Warude and Singh [62] developed a hybrid approach of segmenting nuclei of WBC including the fuzzy c-means clustering, morphological operations, and gradient-based watershed algorithm. Bhukya et al. [63] used Otsu's thresholding for binarization and the morphological operations to segment the nucleus of leukocyte cells in detecting ALL. Then, SVM is utilized for discriminating fit and injurious cells with the accuracy of 92.7% by using the shape features. Vaghela et al. [64] applied different segmentation methods like watershed transformation, k-means clustering, histogram equalization, and morphological operations to reveal RBCs and immature WBCs. Amidst of these, morphological operations outperform with 97.8% of accuracy.

Nee et al. [65] combinedly used thresholding, morphological operations, and watershed algorithm for segmenting the leukocyte and leukoblast cells. This is helpful to categorize various types of acute leukemia, like M2, M5, and M6 with an accurateness level of 94.5%. Dorini et al. [66] adopted morphological operations powered with scale-space features for separating leukocytes with assured results.

4.1.5 Wavelet Transform-Based Segmentation

Sarode et al. [67] introduced an algorithm to detect leukemia using digital cosine transform that outperforms the previous image segmentation algorithm in terms of accuracy by 8.5%. In Sarode et al. [68] proposal, Haar transform is used to identify leukemia. It achieves 13.24% of segmentation accuracy and 23.32% of precision than the former methods. Lakshmikanth and Abdul Khayum [69] developed an AML detection system by taken up Haar wavelet transform and Daubechies wavelet transform techniques to partition the cells. Then, ANN is used as classifier. Mazalan et al. [70] defined an RBC counting system using Circular Hough Transform (CHT) with the accuracy rate of 91.87%.

4.1.6 Active Contour-Based Segmentation

Gim et al. [71] offered a method of segmenting the nucleus and cytoplasm of WBC based on the shapes by the way of mean shift clustering and active contour-based gradient-vector flow snake method, respectively. Mathur et al. [72] used active contour for segmenting nucleus and cytoplasm in reporting the WBC subtypes. The images are processed in the color space called Hue saturation value to fix the seed points.

4.1.7 Region Growing Based Segmentation

Madhloom et al. [73] introduced a scheme of segmentation with the intent to detect ALL, which comprised of mathematical morphology, watershed method, and seeded region growing influenced by histogram equalization for segmenting the leukemic cells, separating the overlying cells and partitioning the cytoplasm as well as nucleus portions correspondingly. This scheme is 96% accurate in parting the leukemic cells and 94% accurate in cytoplasm and nuclei segmentation. Gomez et al. [74] proposed a novel instance-based seeded region growing algorithm. It has been compared with the seeded region growing, and auto-threshold algorithms. Among these, the projected method has shown better results. Abd Halim et al. [75] implemented a method to discover and to discriminate AML and ALL cells. For that, the acquired RGB figure is transformed into HSI color space, *s*-component is selected to attain the threshold value, and then the region growing is applied to segment the regions.

4.1.8 Segmentation Based on Other Mathematical and Statistical Methods

Biji and Hariharan [76] worked on the research identification of WBC using heuristic-based Electromagnetism like Optimization (EMO) method acting as a circle detecting operator. In [77], Gibson et al. derived a sample size formula to estimate the dissimilarities in segmentation accuracy of two algorithms. Raje and Rangole [78] proposed a segmentation scheme in which statistical values like mean and standard deviation are used to partition WBC from erythrocytes and platelets. Geometrical features of WBC are used in classifying regular and blast cells. Nikitaev et al. [79] examined the color components based on texture features of the nucleus of WBCs present in bone marrow images for separating the blast cells and lymphocytes. A mathematical procedure is used to show that component (V) color difference signal is most important in YUV color model.

Reta et al. [80] introduced a contemporary application in classifying the five different types of ALL cells. In this, the Markov random field algorithm is applied to get the color and texture information for acquiring the cytoplasm and nucleus regions of cells. It gains the accuracy of 90%. Mohammed et al. [81] introduced a segmentation technique for separating the nucleus by converting RGB color space into the color space C-Y and using the brightness element with 98.38% accuracy in segmentation.

4.2 Machine Learning

Leukemic malignancies are classified from healthy leukocytes with the aid of machine learning approaches. Decision trees, neural networks, support vector machines, and K-nearest neighboring methods are playing the significant role of classification in machine learning realm.

4.2.1 Classification Based on Decision Trees

In [82], Serbouti et al. proposed a work to analyze blood malignancies and used Classification And Regression Trees (CART) tool for classification. It gives a fine differentiation between ALL with Acute Undifferentiated Leukemia (AUL). Ko et al. [83] proved that nucleus is enough to classify WBC and they also proved that random forest is a sensible classifier to classify the types of WBC. This is another variation of decision tree kind of machine learning.

4.2.2 Classification with Neural Networks and Fuzzy Logic

In [84], a blood cell diagnosing and classification system was proposed by Kim et al. for investigating and diagnosing blood cells. In their proposal, a neural network-based model is utilized for proficient classification of RBCs and WBCs. Jati et al. [85] proposed an involuntary partitioning method to separate the nucleus of leukocytes on regular and noisiest environmental images. This task is executed with the application of exponential intuitionistic fuzzy deviation method. This method gives 98.52% of segmentation accuracy in the regular noiseless environment and 93.90 and 94.93% accuracy rates for Speckle and Gaussian noises, respectively. The region that comes under Operator Characteristic Curve (ROC) is about 0.9514. Asl and Zarandi [86] introduced a type-2 fuzzy expert system to diagnose leukemia using the Mamdani-style inference that provides 94% of classification accuracy.

Fatma and Sharma [87] designed for the cause of search out and classifying leukemia through extracting the features from blood smear images. Taken features are given as input to the neural (semantic networks) network in sorting out the leukemia categories. Colantonio et al. proposal is [88] segmentation of nuclei of leukocytes in finding the malignancy through the application of the fuzzy c-means method in the first stage and ANN in the second stage of uncovering subcomponents of blood cells. Singh et al. [89] introduced a procedure to detect leukemia that uses SIFT method to take out the features; extracted features are modified to get maximum efficiency with Bacterial Foraging Optimization (BFO); and the neural network is employed in classifying the feature with the accuracy of 97.22%. Garg and Kaur [90] exercised k-means technique for segmentation and backpropagation neural networks for classifying leukemic and non-leukemic cells.

4.2.3 Classification with Support Vector Machines

Ramose et al. [91] presented a method to segment and classify leukocytes influenced by cell appearance and image quality. In this proposal, pair-wise Support Vector

Machines (SVMs) classification is used to distinguish the cells of different types. Tai et al. [92] introduced a segmentation and classification scheme to distinguish various categories of blood cells. Classification is accomplished hierarchically started from the geometric features of nucleus and cytoplasm through multi-class SVM.

The procedure proposed by Rawat et al. [93] distinguishes ALL cells and healthy WBCs using shape features and GLCM. Auto SVM classifies the extracted features. It shows the accuracy in classification as 86.7% for the nucleus, 72.4% for cytoplasm, and together it shows the accuracy of 89.8%. Their research reports that in differentiating normally matured and immature lymphocytes, the nucleus shape is more crucial than cytoplasms. Mohapatra and Patra [94] have used the features Hausdorff dimension and contour signature to make out the shape irregularities in the boundary of leukocytes nuclei. SVM is used as the classifier for classifying the leukemic cells. Bigorra et al. [95] presented a technique of distinguishing Reactive Lymphoblastic Cells (RFCs), lymphoblastic, and myeloid blast cells from blood cell images with SVM. This method provides the classification accuracy of 90.1% with 220 test images.

Patil and Raskar [96] suggested a scheme to detect blast cells by means of segmenting the nuclei using Otsu's threshold, and SVM is applied for classification. Moradi Amin et al. [97] proposed a system to categorize the four types of leukemia by SVM. Meera and Mathew [98] have segmented the nuclei region of WBC using fuzzy local information-based c-means clustering, features extracted using GLCM combined with fractal dimensions, and then classified with SVM. Pan et al. [99] applied mean shift scheme and learning by training kind of trained SVM to partition the nucleus of WBCs. Mohapatra et al. [100] proposed a color-based clustering scheme of segmentation that includes *k*-medoid, k-means, Fuzzy Possibilistic C-Means (FPCM) along with Gustafson Kessel to separate nuclei of WBCs in perceiving ALL cells and after that SVM is employed as a classifier. Ravikumar and Shanmugam [101] proposed an algorithm named relevance vector machine a change of SVM for classifying the leukocytes and it is efficient of 91%. James and Nair [102] have done segmentation of WBC in revealing AML using k-means algorithm. In this work, healthy and cancerous cells are classified with SVM.

4.2.4 *K*-Nearest Neighbor Classification

Asadi et al. [103] offered a holographic system to identify and classify an unknown leukemia cell. Feature extraction has been done with Zernike moments. To classify leukemia, KNN and Minimum Mean Distance (MMD) methods are applied. Joshi et al. [104] proposed an automatic segmentation combined with KNN classifier to classify normal lymphocytes and lymphoblast cells with the classification accuracy rate of 93%. Abdeldaim et al. [105] investigated a method to identify ALL by converting the RGB image to CMYK color space, threshold by Zack technique, classification is done using different classifiers, and finally *K*NN classifier achieved good classification accuracy. Di Ruberto et al. [106] presented a new technique combining nearest neighbor along with SVM classifiers in segmenting various components of blood cytoplasm and nucleus of WBC, RBCs, and background. This scheme ensures accuracy in the segmentation about 99% with the dataset ALL-IDB.

5 Research Issues

To substantiate the researchers of medical image processing, several standard datasets of images of different types of diseases are made freely available. In these years, lots of efforts have been put on analyzing and finding new methodologies in recognition and classification of leukemia by using the images in benchmark leukemia datasets such as ALL-IDB 1, ALL-IDB 2, HistologyDS2828, LYMPH, BloodSeg, leukocytes, LISC, and MLL_leukemia. Henceforth, researchers have to explore new datasets of images to show accuracies. All the investigations have not been made on the same dataset. Resultantly, comparisons about accuracy and efficiency of algorithms cannot be made easily. Though the developed methods work well, eventually it is necessary to get substantiation from a hematologist or pathologist. For the cause of saving lives of human, the algorithms in medical image processing must be designed with care and tested on datasets of a large number of images. In view of the fact that medical images are not in easy access, complications are faced in experimenting and showing accuracy with a huge real-time dataset.

6 Conclusion

Research activities accomplished in medical image processing make an effort to reduce the percentage of mortality among living beings. With this insight, an analysis and study over the area of diagnosis of blood cells and leukemia detection are made by referring the research works from the year 1991 to till date. It is observed that by applying multifarious combination of image processing and machine learning techniques on microscopic blood smear images, the leukemia detection and classification can be done through segmenting the cytoplasm and the nuclei of WBC. Some of the research activities have been done on classifying the four major subtypes of leukemia. Derived methods are evaluated and verified by means of extracting the texture, contour, and color features. Though sufficient benchmark datasets of microscopic images of blood cells exist, all the studies have not been undergone with the standard identical datasets of same size and resolution. As a result, the proposed methods cannot be easily compared with each other by means of accuracy and efficiency. Therefore, it is obvious that improvements could be shown by the way of finding out innovative segmentation and classification methodologies with new datasets, and it is ought to focus on categorizing the types of leukemia. In our future work, better segmentation algorithm is going to be proposed with improved accuracy in classification. The majority of the research applications have been implemented with the tools MATLAB, Python, and LabVIEW.

References

1. Cseke I (1992) A fast segmentation scheme for white blood cell images. IEEE conference
2. Liao Q, Deng Y (2002) An accurate segmentation method for white blood cell images. IEEE
3. Prasad B, Choi J-SI, Badawy W (2006) A high throughput screening algorithm for leukemia cells. In: IEEE Canadian conference on electrical and computer engineering, Ottawa, pp 2094–2097
4. Prasad B, Badawy W (2007) High throughput algorithm for leukemia cell population statistics on a hemocytometer. In: IEEE biomedical circuits and systems conference, pp 27–30
5. Biswas S, Ghoshal D (2016) Blood cell detection using thresholding estimation based watershed transformation with Sobel filter in frequency domain. In: Twelfth international multi-conference on information processing (IMCIP-2016) (Elsevier)
6. Kit CY, Tomari R, Nurshazwani W, Zakaria W, Othman N, Safuan SNM, Yi JAJ, Sheng NTC (2017) Mobile based automated complete blood count (Auto-CBC) analysis system from blood smeared image. Int J Electr Comput Eng (IJECE) 7(6):3020–3029. ISSN 2088-8708
7. Scotti F (2006) Robust segmentation and measurements techniques of white cells in blood microscope images. In: Proceedings of the IEEE instrumentation and measurement technology conference, pp 43–48
8. Rezatofighi SH, Zoroofi RA, Sharifian R, Soltanian-Zadeh H (2008) Segmentation of nucleus and cytoplasm of white blood cells using gram-schmid orthogonalization of deformable models. In: 9th international conference on signal processing, IEEE Explore
9. Bergen T, Steckhan D, Wittenberg T, Zerfass T (2008) Segmentation of leukocytes and erythrocytes in blood smear images. In: 30th annual international IEEE EMBS conference Vancouver, British Columbia, Canada
10. Rezatofighi SH, Soltanian-Zadeh H, Sharifian R, Zoroofi RA (2009) A new approach to white blood cell nucleus segmentation based on gram-schmidt orthogonalization. In: International conference on digital image processing, IEEE
11. Sadeghian F, Seman Z, Ramli AR, Kahar BHA, Saripan M-I (2009) A framework for white blood cell segmentation in microscopic blood images using digital image processing. Biological Procedures Online, Springer, New York, pp 196–206
12. Nor Hazlyna H, Mashor MY, Mokhtar NR, Aimi Salihah AN, Hassan R, Raof RAA, Osman MK (2010) Comparison of acute leukemia image segmentation using HSI and RGB colorspace. In: 10th international conference on information sciences signal processing and their applications (ISSPA), pp 749–752
13. Mohamed M, Far M (2012) An enhanced threshold based technique for white blood cells nuclei automatic segmentation. In: 14th international conference on e-health networking, applications and service
14. Di Ruberto C, Putzu L (2014) Accurate blood cells segmentation. In: 2014 tenth international conference on signal-image technology and internet-based systems
15. Ahasan R, Ratul AU, Bakibillah ASM (2016) White blood cells nucleus segmentation from microscopic images of strained peripheral blood film during leukemia and normal condition. In: 5th international conference on informatics, electronics and vision (ICIEV)
16. Lina, Chris A, Mulyavan B, Dharmawan AB (2016) Leukocyte detection using image stitching and color overlapping windows. Int J Comput Electr Autom Control Inf Eng 10(5) (World Academy of Science, Engineering and Technology)
17. Ananthi VP, Balasubramaniam P (2016) A new thresholding technique based on fuzzy set as an application to leukocyte nucleus segmentation. Comput Methods Programs Biomed
18. Sandhu RK (2017) Comparative study of various techniques for leukemia detection. Int J Comput Sci Eng (IJCSE) 9(5). ISSN 0975-3397
19. Putzu L, Di Ruberto C (2013) White blood cells identification and counting from microscopic blood image. Int J Med Health Biomed Bioeng Pharm Eng 7(1) (World Academy of Science, Engineering and Technology)
20. Gosh M, Das D, Chakraborty C, Ray AK (2010) Automated leukocyte recognition using fuzzy divergence. Micron 41:840–846 (Elsevier Ltd.)

21. Khobragade S, Mor DD, Patil CY (2015) Detection of leukemia in microscopic white blood cell images. Int Conf Image Process (ICIP)
22. Abbas N, Mohamad D (2014) Automatic color nuclei segmentation of leukocytes for acute leukemia. Res J Appl Sci Eng Technol 7(14):2987–2993
23. Khashman A, Al-Zgoul E (2010) Image segmentation of blood cells in leukemia patients. Recent Adv Comput Eng Appl 104–109 (ACM Digital Library)
24. Abbas N, Mohamad D, Abdullah AH, Saba T, Al-Rodhaan M, Al-Dhelaan A (2015) Nuclei segmentation of leukocytes in blood smear digital images. Pakistan J Pharm Sci 28(5):1801–1806
25. Li Y, Zhu R, Mi L, Cao Y, Yao D (2016) Segmentation of white blood cell from acute lymphoblastic leukemia images using dual-threshold method. Comput Math Methods Med 2016 (Hindawi)
26. Nilsson B, Heyden A (2001) Segmentation of dense leukocyte clusters. In: IEEE workshop on mathematical methods in biomedical image analysis (MMBIA01), pp 221–227
27. Jiang K, Liao Q-M, Dai Mach S-Y (2003) A novel white blood cell segmentation using scale-space filtering and watershed clustering (vol 5, pp 2820–2825). In: Proceedings of the second international conference on machine learning and cybernetics, Xi'an, IEEE, pp 2–5
28. Shankar V, Deshpande MM, Chaitra N, Aditi S (2016) Automatic detection of acute lymphoblastic leukemia using image processing. In: IEEE international conference on advances in computer applications
29. Ghane N, Vard A, Talebi A, Nematollahy P (2017) Segmentation of white blood cells from microscopic images using a novel combination of K-means clustering and modified watershed algorithm. J Med Signals Sensors 7(2):92–101
30. Belekar SJ, Chougule SR (2015) WBC segmentation using morphological operation and SMMT operator—a review. Int J Innov Res Comput Commun Eng 3(1). ISSN 2320-9801
31. Liu Z, Liu J, Xiao X, Yuan H, Li X, Chang J, Zheng C (2015) Segmentation of white blood cells through nucleus mark watershed operations and mean shift clustering. Sensors 15:22561–22586
32. Sinha N, Ramakrishnan AG (2003) Automation of differential blood count, IEEE
33. Basima CT, Panicker JR (2016) Enhance leukocyte classification for leukemia detection. IEEE
34. Foran DJ, Comaniciu D, Meer P, Goodell LA (2000) Computer-assisted discrimination among malignant lymphomas and leukemia using immunophenotyping. In: Intelligent image repositories and telemicroscopy. IEEE Transactions on Information Technology in Biomedicine
35. Su J, Liu S, Song J (2017) A segmentation method based on HMRF for the aided diagnosis of acute myeloid leukemia. Comput Methods Programs Biomed. https://doi.org/10.1016/j.cmpb.2017.09.011
36. Laosai J, Chamnongthai K (2014) Acute leukemia classification by using SVM and K-means clustering. In: Proceedings of the international electrical engineering congress, IEEE
37. Kumar R, Srivastava R, Srivastava S (2015) Detection and classification of cancer from microscopic biopsy images using clinically significant and biologically interpretable features. J Med Eng 2015 (Hindawi Publishing Corporation)
38. Neoh SC, Srisukkham W, Zhang L, Todryk S, Greystoke B, Lim CP, Hossain MA, Aslam N (2015) An intelligent decision support system for leukemia diagnosis using microscopic blood images. J Sci Reports
39. Agaian S, Madhukumar M, Chronopoulos AT (2014) Automated screening system for acute myelogenous leukemia detection in blood microscopic images. IEEE Syst J 8(3)
40. Moradi Amin M, Kermani S, Talebi A, Oghli MG (2015) Recognition of acute lymphoblastic leukemia cells in microscopic images using K-means clustering and support vector machines. J Med Signals Sensors 5(1):49–58
41. Hazra T, Kumar M, Tripathy SS (2017) Automatic leukemia detection using image processing technique. Int J Latest Technol Eng Manag Appl Sci (IJLTEMAS) 6(4). ISSN 2278-2540
42. Indira P, Ganesh Babu TR, Vidhya K (2016) Detection of leukemia in blood microscope images. Int J Control Theory Appl 9(5):2147–2151

43. Madhukar M, Agaian S, Chronopoulos AT (2012) Deterministic model for acute myelogenous leukemia classification. In: 2012 IEEE international conference on systems, man, and cybernetics, COEX, Seoul, Korea, 14–17 Oct 2012
44. Soltanzadeh R, Rabbani H, Talebi A (2012) Extraction of nucleolus candidate zone in white blood cells of peripheral blood smear images using curvelet transform. Comput Math Methods Med 2012
45. Sajjad M, Khan S, Jan Z, Muhammad K, Hyeonjoon Moon, Kwak JT, Rho S, Baik SW, Mehmood I (2017) Leukocytes classification and segmentation in microscopic blood smear: a resource-aware healthcare service in smart cities. In: Special section on advances of multisensory services and technologies for healthcare in smart cities, vol 5, IEEE
46. Selvaraj S, Kanakaraj B (2015) Naïve Bayesian classifier for acute lymphocytic leukemia detection. ARPN J Eng Appl Sci 10(16). ISSN 1819-6608
47. Sivakumar S, Ramesh S (0215) Automatic white blood cell segmentation using K-means clustering. Int J Sci Eng Res 3(4)
48. Kazemi F, Najafabadi TA, Araabi BN (2016) Automatic recognition of acute myelogenous leukemia in blood microscopic images using K-means clustering and support vector machines. J Med Signals Sens
49. Sharma N, Kinra N (2014) Detecting and counting the no of white blood cells in blood sample images by color based K-means clustering. Int J Electr Electron Eng 1(3). ISSN 1694–2310
50. Harun NH, Absdul Nasir AS, Mashor MY, Hassan R (2015) Unsupervised segmentation technique for acute leukemia cells using clustering algorithms. Int J Comput Inf Eng 9(1) (World Academy of Science, Engineering and Technology)
51. Piuri V, Scotti F (2004) Morphological classification of blood leukocytes by microscopic images. In: IEEE international conference on computational intelligence for measurement systems and applications
52. Theera-Umpon N, Dhompongsa S (2007) Morphological granulometric features of nucleus in automatic bone marrow white blood cell classification. IEEE Trans Inf Technol Biomed 11(3):353–359
53. Vogado LHS, Rodrigo de M. S. Veras, Andrade AR, Romuere R. V. e Silva, Flavio H. D. de Araujo, de Medeiros FNS (2016) Unsupervised leukemia cells segmentation based on multi-space color channels. In: IEEE international symposium on multimedia
54. Bouzid-Daho A, Boughazi M, Tanouast C (2017) Algorithmic processing to aid in leukemia detection. Med Technol J 1(1):10–11
55. Bhattacharjee R, Saini LM (2015) Robust technique for the detection of acute lymphoblastic leukemia. In: Communication and information technology conference (PCITC) Siksha 'O' Anusandhan University, Bhubaneswar, India, IEEE Power
56. Scotti F (2005) Automatic morphological analysis for acute leukemia identification in peripheral blood microscopic images. In: IEEE international conference on computational Intelligence for measurement systems and Applications
57. Vishwanathan P (2015) Fuzzy C means detection of leukemia based on morphological contour segmentation. Procedia Comput Sci 58:84–90 (Elsevier)
58. Gumble P, Rode SV (2017) Study and analysis of acute lymphoblastic leukemia blood cells using image processing. Int J Innov Res Comput Commun Eng 5(1). ISSN (online): 2320–9801
59. Polyakov EV, Nikitaev VG (2017) A method for estimating the accuracy of measurements of optical characteristics of the nuclei of blood cells in the diagnosis of acute leukemia. J Phys Conf Ser 784:012042
60. Grimwade LF, Fuller KA, Erber WN (2016) Applications of imaging flow cytometry in the diagnostic assessment of acute leukemia. Elsevier
61. Singhal V, Singh P (2014) Local binary pattern for automatic detection of acute lymphoblastic leukemia. IEEE
62. Warude D, Singh R (2016) Automatic detection method of leukemia by using segmentation method. Int J Adv Res Comput Commun Eng 5(3)

63. Bhukya R, Prasanth B, Sasank Vihari V, Ajay Y (2017) Detection of acute lymphoblastic using microscopic images of blood. Int J Adv Appl Sci 4(8):74–78
64. Vaghela HP, Modi H, Pandya M, Potdar MB (2015) Leukemia detection using digital image processing techniques. Int J Appl Inf Syst (IJAIS) 10(1). ISSN: 2249-0868
65. Nee LH, Mashor MY, Hassan R (2012) White blood cell segmentation for acute leukemia bone marrow images, In: 2012 international conference on biomedical engineering (ICoBE), IEEE
66. Dorini LB, Minetto R, Leite NJ (2007) White blood cell segmentation using morphological operators and scale-space analysis. Comput Graph Image Process. ISSN 1530-1834 (IEEEXplore)
67. Sarode TK, Thakkar BK, Purandare SJ, Gupta VM (2016) Cancerous cell detection in bone marrow smear. Int J Comput Appl
68. Sarode TK, Thakkar B, Purandare SJ, Gupta VM (2016) Cancerous cell detection in bone marrow smear using Haar transform. In: International conference and workshop on electronics & telecommunication engineering (ICWET 2016), IEEE
69. Lakshmikanth BK, Abdul khayum P (2017) Acute myelogenous leukemia detection in blood microscopic images using different wavelet family techniques. Int J Eng Manag Res 7(4):174–182
70. Mazalan SM, Mahmood NH, Razak MAA (2013) Automated red blood cells counting in peripheral blood smear image using circular hough transform. In: 2013 first international conference on artificial intelligence, modeling & simulation, IEEEXplore
71. Gim J-W, Park J, Lee J-H, Ko BC, Nam J-Y (2011) A novel framework for white blood cell segmentation based on stepwise rules and morphological features. In: Proceedings of SPIE-IS&T 7877, image processing machine vision applications, San Francisco, pp 1–6
72. Mathur A, Tripathi AS, Kuse M (2012) Scalable system for classification of white blood cells from leishman stained blood stain images. J Pathol Inform (HIMA workshop at MICCAI, Nice, France)
73. Madhloom HT, Kareem SA, Ariffin H (2015) Computer-aided acute leukemia blast cells segmentation in peripheral blood images. J VibroEng 17(8):4517–4532
74. Gomez O, Gonzalez JA, Morales EF (2007) Image segmentation using automatic seeded region growing and instance-based learning. In: Progress in pattern recognition, image analysis and applications, CIARP 2007, Springer, Berlin, Heidelberg
75. Abd Halim NH, Mashor MY, Abdul Nasir AS, Mokhtar NR, Rosline H (2011) Nucleus segmentation technique for acute leukemia. In: 2011 IEEE 7th international colloquium on signal processing and its applications
76. Biji G, Hariharan S (2017) An efficient peripheral blood smear image analysis technique for leukemia detection. In: International conference on I-SMAC (IOT in Social, Mobile, Analytics, and Cloud), I-SMAC, IEEE
77. Gibson E, Hu Y, Huisman HJ, Barratt DC (2017) Designing image segmentation studies: statistical power, sample size, and reference standard quality. Med Image Anal 42:44–59 (Elsevier)
78. Raje C, Rangole J (2014) Detection of leukemia in microscopic images using image processing. In: International conference on communication and signal processing
79. Nikitaev VG, Pronichev AN, Polyakov EV, Dmitrieva VV, Tupitsyn NN, Frenkel MA, Mozhenkoa AV (2017) The influence of physical factors on recognizing blood cells in the computer microscopy systems of acute leukemia diagnosis. J Phys Conf Series 798:012128
80. Reta C, Altamirano L, Gonzalez JA, Diaz-Hernandez R, Peregrina H, Olmos I, Alonso JE, Lobato R (2015) Segmentation and classification of bone marrow cells images using contextual information for medical diagnosis of acute leukemias. PLOS One 10(6)
81. Mohammed R, Nomir O, Khalifa I (2014) Segmentation of acute lymphoblastic leukemia using C-Y color space. Int J Adv Comput Sci Appl (IJACSA) 5(11)
82. Serbouti S, Duhamel A, Harms H, Gunzer U, Aus HM, Mary JY, Beuscart R (1991) Image segmentation and classification methods to detect leukemias. In: Annual international conference of IEEE engineering in medicine and biology society

83. Ko BC, Gim JW, Nam JY (2011) Cell image classification based on ensemble features and random forest. Electron Lett 47(11):638–639
84. Kim KS, Kim PK, Song JJ, Park YC (2002) Analyzing blood cell images do distinguish its abnormalities. In: ACM international conference on multimedia, Los Angeles, CA, USA, pp 395–397
85. Jati A, Singh G, Mukherjee R, Gosh M, Konar A, Chakraborty C, Nagar AK (2014) Automatic leukocyte nucleus segmentation by intuitionistic fuzzy divergence based thresholding. Micron 58:55–65 (Elsevier Ltd.)
86. Asl AAS, Zarandi MHF (2018) A type-2 fuzzy expert system for diagnosis of leukemia. In: Fuzzy logic in intelligent system design, advances in intelligent systems and computing, vol 648. Springer International Journal AG (in press)
87. Fatma M, Sharma J (2014) Identification and classification of acute leukemia using neural networks. In: 2014 international conference on medical imaging, m-health and emerging communication systems (MedCom), IEEE
88. Colantonio S, Gurevich IB, Salvetti O (2008) Automatic fuzzy-neural based segmentation of microscopic cell images. Int J Signal Image Syst Eng 1(1)
89. Singh G, Bathla G, Kaur S: (2016) Design of new architecture to detect leukemia cancer from medical images. Int J Appl Eng Res 11(10):7087–7094. ISSN 0973-4562 (Research India Publications)
90. Garg J, Kaur D (2016) Automated blood cancer detection (leukemia) using artificial intelligence by ACO algorithm with BPNN classifier under soft computing 4(3). ISSN 2321-2632
91. Ramoser H, Lauria V, Bischof H, Ecker R (2008) Leukocyte segmentation and SVM classification in blood smear images. Graph Vis 17(1):187–200
92. Tai W-L, Hu R-M, Hsiao HCW, Chen R-M, Tsai JJP (2011) Blood cell image classification based on hierarchical SVM. Department of Biomedical Informatics, Sia University, Taiwan, IEEE
93. Rawat J, Singh A, Bhadauria HS, Virmani J (2015) Computer aided diagnostic system for detection of leukemia using microscopic images. Procedia Comput Sci 70:748–756 (Elsevier)
94. Mohapatra S, Patra D (2010) Automatic cell nucleus segmentation and acute leukemia detection in blood microscopic images. In: International conference on systems in medicine and biology, IIT Kharagpur India
95. Bigorra L, Merino A, Alferez S, Rodellar J (2016) Feature analysis and automatic identification of leukemic lineage blast cells and reactive lymphoid cells from peripheral blood cell images. J Clin Lab Anal 00:1–9
96. Patil TG, Raskar VB (2015) Automated leukemia detection by using contour signature method. Int J Adv Found Res Comput (IJAFRC) 2(6). ISSN 2348-4853
97. MoradiAmin M, Memari A, Samadzadehaghdam N, Kermani S, Talebi A (2016) Computer aided detection and classification of acute lymphoblastic leukemia cell subtypes based on microscopic image analysis. Microsc Res Techn 908–916 (Wiley Periodicals, Inc.)
98. Meera V, Mathew SA (2014) Fuzzy local information C means clustering for acute myelogenous leukemia image segmentation. Int J Innov Res Sci, Eng Technol 3(5). ISSN 2319-8753
99. Pan C, Lu H, Cao F (2009) Segmentation of blood and bone marrow cell images via learning by sampling. In: Emerging intelligent computing technology and applications, ICIC 2009, vol 5754. Springer, pp 336–345
100. Mohapatra S, Patra D, Satpathy S (2012) Unsupervised blood microscopic image segmentation and leukemia detection using color based clustering. Int J Comput Inf Syst Ind Manag Appl 4:477-485. ISSN 2150-7988
101. Ravikumar S, Shanmugam A (2014) WBC image segmentation and classification using RVM. Appli Math Sci 8(45):2227–2237 (HIKARI Ltd.)
102. James J, Nair KN (2013) Automated acute myelogenous leukemia detection in blood microscopic image. Int J Sci Res (IJSR). ISSN 2319-7064
103. Asadi MR, Vahedi A, Amindavar H (2006) Leukemia cell recognition with zernike moments of holographic images. In: IEEE proceedings of 7th nordic signal processing symposium, Iceland, pp 214–217

104. Joshi MD, Karode AH, Suralkar SR (2013) White blood cells segmentation and classification to detect acute leukemia. Int J Emerg Trends Technol Comput Sci
105. Abdeldaim AM, Sahlol AT, Elhoseny M, Hassanien AE (2018) Computer-aided acute lymphoblastic leukemia diagnosis system based on image analysis. In: Advances in soft computing and machine learning in image processing. Springer International Publishing AG (in press)
106. Di Ruberto C, Loddo A, Putzu L (2015) A multiple classifier learning by sampling system for white blood cells segmentation. In: International conference on computer analysis of images and patterns, CAIP 2015, vol 9257. Springer, pp 415–425

D. Umamaheswari has completed her M.Sc. and M.Phil. degrees in Computer Science and published five papers in international journals. She has 12 years of teaching experience and qualified in State Level Eligibility Test (SLET). At present, she is working in Vidyasagar College of Arts and Science, Udumalepet, Tamilnadu. Currently, she is pursuing Ph.D. (Part-time) in Computer Science, Government Arts College, Udumalpet, Tamilnadu, India. Her area of interest is image processing.

Dr. S. Geetha has completed her M.Sc., M.Phil., and Ph.D. degrees in Computer Science. She has published various papers in international journals of repute. Her area of specialization is soft computing. She has 20 years of experience in PSG College of Technology, Coimbatore, Tamilnadu. Presently, she is working at LRG Government Arts College for Women, Tiruppur, Tamilnadu, India.

Detection of Gaze Direction for Human–Computer Interaction

G. Merlin Sheeba and Abitha Memala

Abstract Eye guide is an assistive specialized apparatus intended for the incapacitated or physically disabled individuals who were not able to move parts of their body, especially people whose communications are limited only to eye movements. The prototype consists of a camera and a computer. The system recognizes gazes in four directions and performs required user actions in related directions. The detected eye direction can then be used to control the applications. The facial regions which form the images are extracted using the skin color model and connected-component analysis. When the eye regions are detected, the tracking is performed. The system models consist of image processing, face detector, face tracker, and eyeblink detection. The eye guide system potentially helps as a computer input control device for the disabled people with severe paralysis.

1 Introduction

Human–Computer Interaction (HCI) is the study of how people interact with computers and to what extent computer is not developed for successful interaction with human beings. The goals of HCI are to produce usable and safe systems, as well as functional systems. The human–computer interface can be portrayed as the purpose of correspondence between the human client and the Personal Computer (PC) [1–3]. The surge of information between the human and PC is described as the hover of correspondence. The circle of cooperation has a few angles to it, including the following:

- Visual Based: The visual-based human PC between activities likely the most across the board region in HCI look into.
- Audio Based: The sound-based collaboration between a PC and a human is another critical territory in HCI frameworks. This territory manages data gained by various sound signs.

G. M. Sheeba (✉) · A. Memala
Sathyabama Institute of Science and Technology, Chennai, TamilNadu, India
e-mail: merlinsheebu@gmail.com

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_164

- Task Environment: The conditions and objectives set upon the client.
- Machine Environment: The environment that the PC is associated with.
- Output: The stream of data that begins in the machine environment.
- Feedback: Loops through the interfaces that assess, direct, and affirm forms as they go from the human through the interface to the PC and back.

2 System Description

2.1 System Model

The components of the system shown in Fig. 1 are explained as follows: The video input is converted into frames and the image is processed further.

2.1.1 Conversion of Grayscale

Each frame is processed into a grayscale image. The grayscale is a picture in which the estimation of pixel is a solitary example, i.e., it conveys just force data. The dark scale pictures are likewise called as highly contrasting pictures. Here, the weakest part will be in dark and most grounded part will be white as shown in Fig. 2.

2.1.2 Scanning

It is the process of translating the images or photographs into a digital form that can be recognized by a computer.

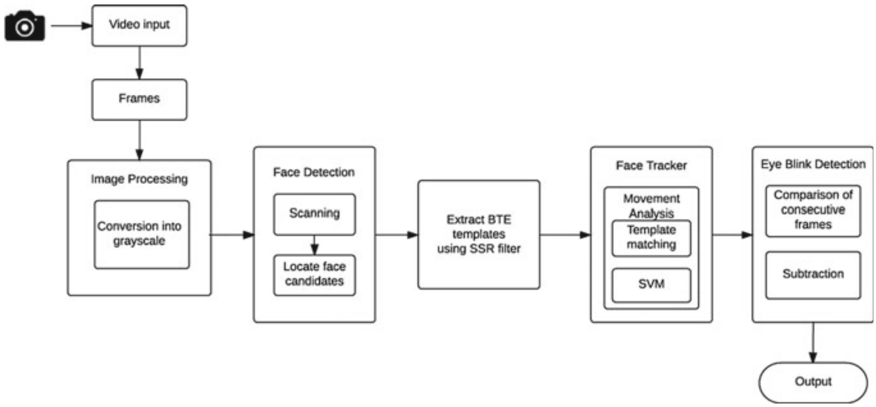


Fig. 1 The human–computer interaction system model

2.1.3 Six-Segmented Rectangular Filter (SSR)

A 2×3 rectangle is considered to examine on the selected frame. This rectangle is fragmented into six sections as demonstrated as follows.

The SSR channel is utilized to distinguish between the eyes in light of two qualities of face geometry.

- The nose zone (Sn) is brighter than the privilege and left eye region (Ser and Sel, separately) as appeared in Fig. 2b, where

$$\mathrm{Sn} = Sb2 + Sb5 \tag{1}$$

$$\mathrm{Ser} = Sb1 + Sb4 \tag{2}$$

$$\mathrm{Sel} = Sb3 + Sb6 \tag{3}$$

At that point,

$$\mathrm{Sn} > \mathrm{Ser} \tag{4}$$

$$\mathrm{Sn} > \mathrm{Sel} \tag{5}$$

- The eye range (both eyes and eyebrows) (Se) is generally darker than the cheekbone region (counting nose) (Sc) as appeared in Fig. 2c, where

$$\mathrm{Se} = Sb1 + Sb2 + Sb3 \tag{6}$$

$$\mathrm{Sc} = Sb4 + Sb5 + Sb6 \tag{7}$$

At that point,

$$\mathrm{Se} < \mathrm{Sc} \tag{8}$$

Whenever expressions (4), (5), and (8) are altogether fulfilled, the focal point of the rectangle can be a contender for between the eyes.

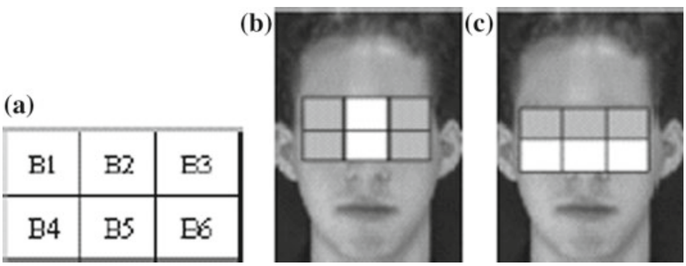


Fig. 2 SSR filter

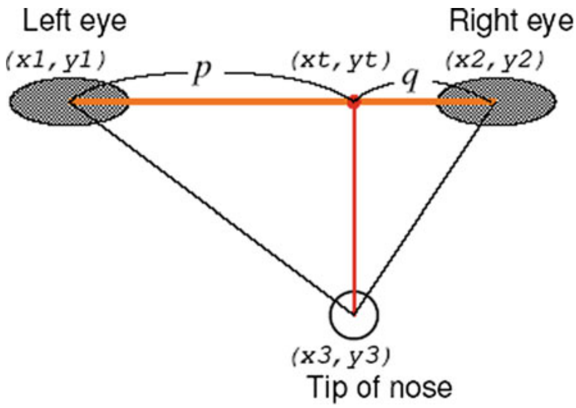


Fig. 3 Determination of BTE

2.1.4 Tracking

The raised shaped elements in the facial features do not have unique property. Hence, using skin shading, outline subtraction, and learning of the component positions, the initial step is to ascertain the region for nearby pursuit of the nose highlight. The second step comprises checking the nearby hunt territory and discovering pixel "*u*". Before continuing to this progression, the picture is reprocessed with the Gaussian channel to smooth the deformities of pictures brought about by low nature of the cameras [4]. The last stride is to refine the position of the best match, utilizing the confirmation-based convolution channel. The acquired match of the raised shape highlight is both strong (to turn and scale) and exact (processed with sub-pixel exactness).

2.1.5 Between the Eye (BTE)

The identification of BTE depends on the property of the picture quality specifically the region on face. The BTE is the seat point on the hyperbolic surface [5]. A rotationally invariant channel could be formulated for identifying the BTE region as shown in Fig. 3.

The basic BTE region on human face looks like a hyperbolic surface. For example, the hyperbolic surface focal and seat points are indicated in the facial image in Fig. 4.

2.1.6 Template Matching

It is a method in computerized picture preparing for discovering little parts of a picture which coordinate a format picture. It can be utilized as a part of assembling as a piece of value control, an approach to explore and to distinguish the edge of image [6].

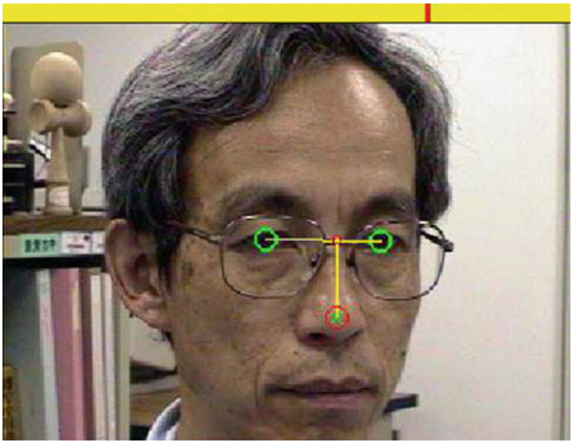


Fig. 4 Indication of the focal and seat points in a facial image

2.1.7 Support Vector Machine

Support vector machines (SVMs) are an arrangement of related regulated learning techniques utilized for order and relapse. SVM takes as information preparing information tests, where each example comprises characteristics and a class mark (positive or negative). The information tests that are nearest to the hyperplane are called support vectors. The hyperplane is characterized by adjusting its separation among positive and negative support vectors with a specific end goal to get the maximal edge of the preparation informational index.

A support vector machine develops a hyperplane or set of hyperplanes in a high or vast dimensional space, which can be utilized for grouping, relapse, or different undertakings [7, 8]. To discover the biggest hyperplane it is required to isolate the positive examples from the negative specimens, which are the two unique classes of information. A decision rule is formulated as below to determine the classes of positive and negative examples:

$$(w \cdot x_i + b) \geq 1, \quad \text{if } y_i = 1 \tag{9}$$

$$(w \cdot x_i + b) \leq -1, \quad \text{if } y_i = -1. \tag{10}$$

2.1.8 Detection of Eyeblink

The idea of the second request change discovery, which permits one to segregate the neighborhood (latest) change in picture, for example, squint of the eyes, from the worldwide (durable) change, for example, the movement of head, is presented. This idea sets the base for planning complete face-worked control frameworks, in which, utilizing the similarity with mouse, "indicating" is performed by nose and "clicking" is performed by twofold squinting of the eyes. Utilizing similarity with

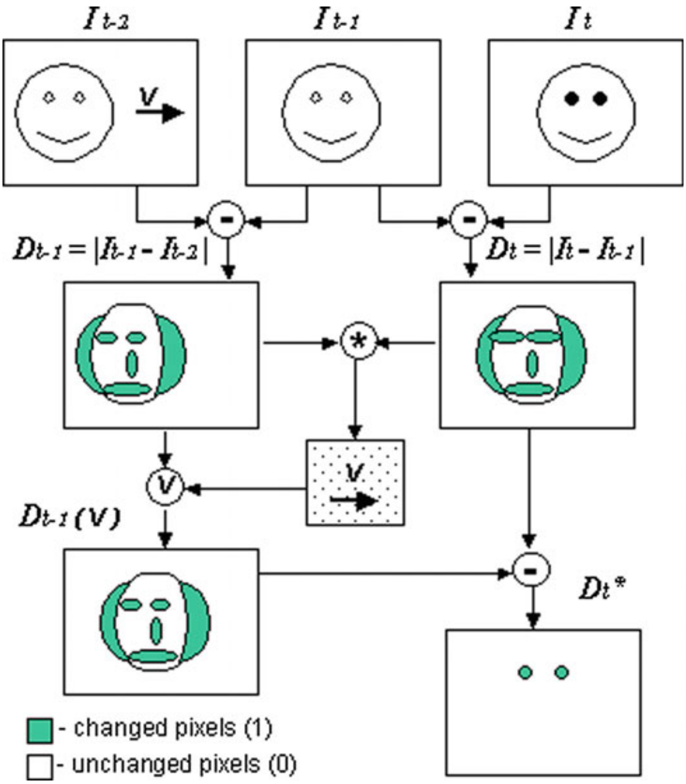


Fig. 5 A flowchart for detecting eyeblinks using second-order change detection

second request subordinates of capacities, the second request changes as a change of a change. The strategy utilized a part of numerical investigation for figuring the second request subordinate of a capacity in light of considering three neighboring focuses, and the second request change in video can be registered by considering three back-to-back edges of the video as shown in Fig. 5.

After comparing the pixels, the head movement has been shifted through the second request change called location methodology; the rest of the pixels are utilized to choose whether a squint has occurred. To begin with, the quantity of the staying changed pixels is analyzed and on the off chance that it is too little, implying that there is no movement, or in the event that it is too huge, implying that the foundation has changed, the eye location procedure is skipped. Something else, the vector quantization system is utilized keeping in mind the end goal to characterize splendid (changed) pixels into two classes, relating to the places of two potential "eyes" as shown in Fig. 6.

Fig. 6 Threshold difference image prior to erosion (left), and same image after erosion (right). Erosion removes noise caused by insignificant motion in the scene

2.1.9 Human Skin Model

A human skin shading model is utilized to choose if shading is a skin or non-skin shading. It is trusted that individuals from various races have same skin shading; yet, they contrast in lighting force of that shading keeping in mind the end goal to locate an appropriate skin shading model; as indicated by the already specified thought, we will utilize the immaculate r and g values which are the R and G estimations of the RGB shading model without shine and they are computed with the accompanying conditions:

$$r = R/(R + G + B) \tag{11}$$

$$g = G/(R + G + B) \tag{12}$$

We move the r and g qualities to the (a, b) shading space with the accompanying conditions:

$$a = r + g/2 \tag{13}$$

$$b = \sqrt{3}/2\,\mathrm{g} \tag{14}$$

The scope of “an” is from 0 to 1, while the scope of “b” is from 0 to $\sqrt{3}/2$.

To digitize the (a, b) space, we round qualities to multiplicands of 0.01.

At last, to discover the skin shading model, we removed 735 skin pixel tests from each of 771 face pictures which were taken from, so the aggregate number of skin pixel tests was 566,685 examples. For each of the examples, we ascertained the “an” and “b” values, and expanded the counter of that esteem (“a” has 101 counters, “b” has 88 counters). The qualities that were most successive are considered as skin pixel values, so subsequent to plotting the “an” and “b” comes about we reasoned the accompanying limits for “an”, and “b” to consider a pixel as a skin pixel:

$$0.49 < \text{“a”} < 0.59 \tag{15}$$

$$0.24 < \text{“a”} < 0.29 \tag{16}$$

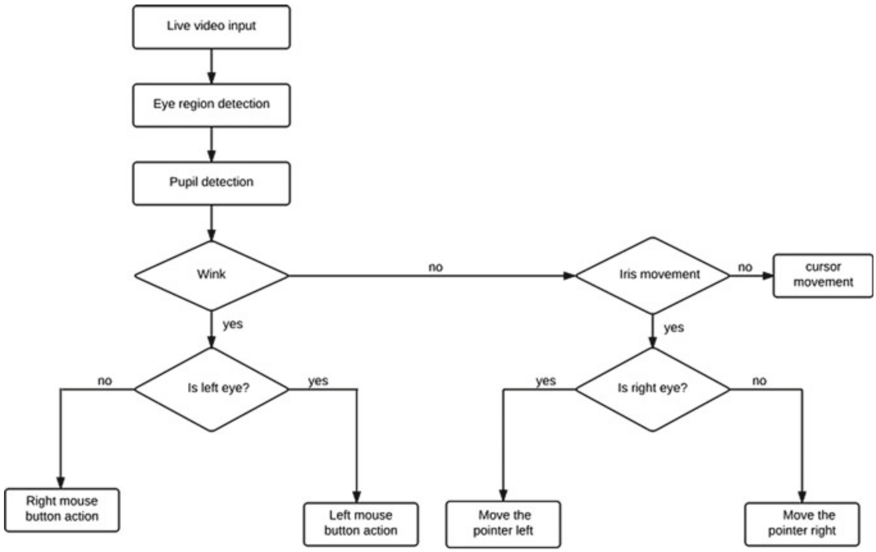


Fig. 7 Flowchart

2.1.10 Model Flow Diagram

The proposed framework utilizes a camcorder to track client's face position in 3D keeping in mind the end goal to change over it to a place of a cursor or another virtual protest in 2D screen [9]. The eyes and nose are found and after that, we begin to track them by breaking down the development of these elements in resulting outlines, utilizing some format coordinating, and heuristics. The framework utilizes the light appeared in the clients' eyes keeping in mind the end goal to find the understudies. The framework then checks if there is any wink in clients' eye. On the off chance that there is any wink recognized, then it checks if the left eye has winked and plays out the left mouse catch activity, else the correct mouse catch activity is performed correspondingly. The framework then checks if there is any development in clients' understudy. On the off chance that any development is identified, then it checks if the left eye has moved and moves the pointer left, else moves the pointer right correspondingly as shown in Fig. 7.

3 Result and Discussions

The eyes and nose are located, and then the systems start to track them by analyzing the movement of these features in subsequent frames, using some template matching, and heuristics. The proposed system uses the light reflection in the user's eyes in order to locate the pupils. The system then checks if there is any wink in user's eye. If there

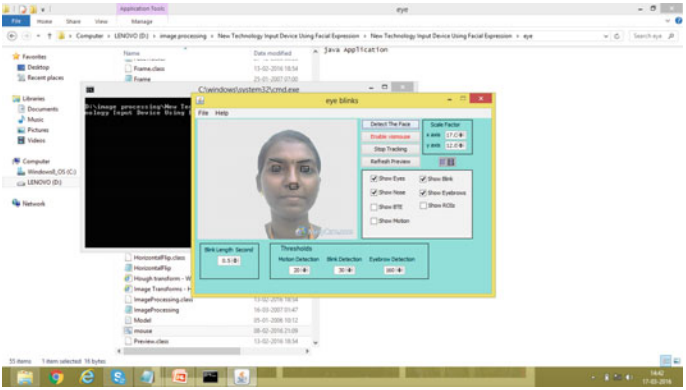


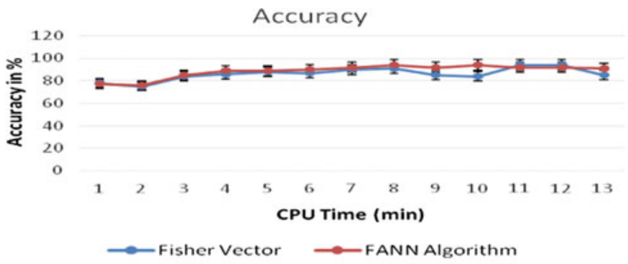
Fig. 8 GUI output

is any wink detected, then it checks if the left eye has winked and performs the left mouse button action, else the right mouse button action is performed correspondingly. The system then checks if there is any movement in user's pupil. If any movement is detected, then it checks if the left eye has moved and moves the pointer left, else moves the pointer right correspondingly. The GUI MATLAB output is shown in Fig. 8.

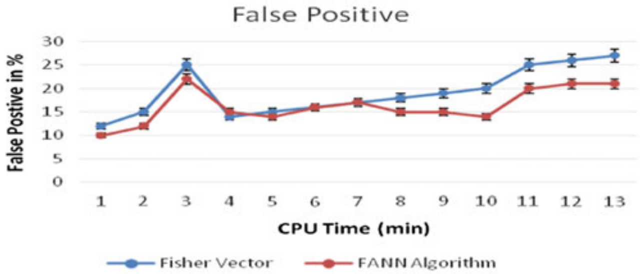
Reproducing the same movements of eyeblink videos is difficult in eye tracking. In this proposed system, extensive experimental measurements are carried out for more samples to evaluate precisely the nose regions and eyebrow movements also. The web camera is assisted with the node-based tracking also [10]. The speed of tracking is high as universal serial bus cameras are fixed to track the blinks.

3.1 Performance Metrics

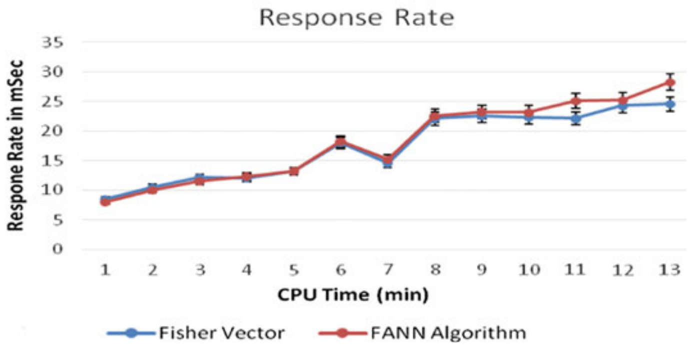
- Accuracy: It refers to closeness of a measured value to standard value of eye recognition. The proposed algorithm is 7% greater than the Fisher vector, which is given in percentage as shown in Fig. 9a.
- False Positive: It refers to the probability of falsely rejecting the null hypothesis for a particular test [11]. It is also calculated as the ratio between the number of negative events wrongly categorized as positive and total number of actual negative events. The proposed algorithm is 22% lesser than the Fisher vector, which is used for eye detection as shown in Fig. 9b.
- Response Rate: It is also known as completion rate to return rate. In this, system calculates the time for eye recognition. The proposed algorithm is 17% high in their speed (ms) than the Fisher vector as shown in Fig. 9c.



(a)Accuracy Vs CPU Time



(b) False Positive Vs CPU Time



(c) Response Rate Vs CPU Time

Fig. 9 Performance evaluation

4 Conclusion

This proposed system based on the face tracking with convex-shape nose feature and eyeblink detection using subtraction methods allows a user the flexibility and convenience of motions. The system serves as an alternative communication method that is suitable for people with severe disabilities. In previous tracking schemes, the disables have to wear special transmitters, sensors, or markers which resulted in discomfort, whereas in this method the eyeblinks use only a web camera and it is completely nonintrusive. The absence of any accessories on the user makes the system easier to configure and therefore more user-friendly in a clinical or academic environment.

Acknowledgements Ethical Compliance Comments
Figure 4 is a facial image taken from UCI repository dataset as an example to indicate the focal and seat points in the face.

References

1. Ebisawa Y (1998) Improved video-based eye-gaze detection method. IEEE Trans Instrum Meas 47(4):948–955
2. Kim KN, Ramakrishna RS (1999) Vision-based eye-gaze tracking for human computer interface. In: 1999 IEEE international conference on systems, man, and cybernetics, 1999. IEEE SMC'99 conference proceedings, vol 2. IEEE, pp 324–329
3. Galante A, Menezes P (2012) A gaze-based interaction system for people with cerebral palsy. Procedia Technol 5:895–902
4. Mahalakshmi E, Sheeba GM (2015) Enhancement of CFA in single sensor camera using laplacian projection technique. Res J Pharm Biol Chem Sci 6(3):1529–1536
5. Magee JJ, Betke M, Gips J, Scott MR, Waber BN (2008) A human–computer interface using symmetry between eyes to detect gaze direction. IEEE Trans Syst, Man, Cybern Part A Syst Humans 38(6):1248–1261
6. Al-Rahayfeh AMER, Faezipour MIAD (2013) Eye tracking and head movement detection: a state-of-art survey. IEEE J Transl Eng Health Med 1:2100212
7. Rantanen V, Vanhala T, Tuisku O, Niemenlehto PH, Verho J, SurakkaV, Juhola M, Lekkala JO (2011) A wearable, wireless gaze tracker with integrated selection command source for human–computer interaction. IEEE Trans Inf Technol BioMedicine 15(5):795–801
8. Carbone A, Martínez F, Pissaloux E, Mazeika D, Velázquez R (2012) On the design of a low cost gaze tracker for interaction. Procedia Technol 3:89–96
9. Zhang L, Vaughan R (2016, October) Optimal robot selection by gaze direction in multi-human multi-robot interaction. In: 2016 IEEE/RSJ international conference on intelligent robots and systems (IROS).IEEE, pp 5077–5083)
10. Miyake T, Asakawa T, Yoshida T, Imamura T, Zhang Z (2009, November). Detection of view direction with a single camera and its application using eye gaze. In: 35th annual conference of IEEE industrial electronics, 2009. IECON'09, pp 2037–2043
11. Merlin Sheeba G (2016) Enhanced wavelet OTSU tracking method for carcinoma cells. Int J Pharm Technol 8(2):11675–11684
12. Santos R, Santos N, Jorge PM, Abrantes A (2014) Eye gaze as a human-computer interface. Procedia Technol 17:376–383
13. Sun L, Liu Z, Sun MT (2015) Real time gaze estimation with a consumer depth camera. Inf Sci 320:346–360

Road Detection by Boundary Extraction Technique and Hough Transform

Namboodiri Sandhya Parameswaran, E. Revathi Achan, V. Subhashree and R. Manjusha

Abstract Visual perception of road images captured by cameras mounted within a vehicle is the main element of an autonomous vehicle system. Road detection plays a vital role in a visual routing system for a self-governing vehicle. Effective detection of roads under varying illumination conditions plays a vital role to prevent majority of the road accidents that occur currently. In the current study, a new method using "boundary extraction" technique along with "Hough transform" is proposed for effective road detection. Here, two different algorithms, one using "Canny edge detection" and "Hough transform" and another using "boundary extraction" technique and "Hough transform" were implemented and tested on the same dataset. The comparison of the results of both the techniques showed that the algorithm using "boundary extraction" technique worked better than that which used "Canny edge" detection technique.

1 Introduction

Massive expansion in road network over the recent years has lead India to achieve the highest growth rate in the world. Our country is facing serious influences on road safety levels. A survey conducted by the Ministry of Road Transport and Highways Transport Research wing, Gov. of India (2015) says that the accident rate has increased by 2–5% from 4,89,400 in 2014 to 5,01,423 in 2015. An automated road

N. S. Parameswaran · E. Revathi Achan · V. Subhashree · R. Manjusha (✉)
Department of Computer Science and Engineering,
Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Coimbatore, India
e-mail: r_manjusha@cb.amrita.edu

N. S. Parameswaran
e-mail: cb.en.p2cvi16004@cb.students.amrita.edu

E. Revathi Achan
e-mail: cb.en.p2cvi16006@cb.students.amrita.edu

V. Subhashree
e-mail: cb.en.p2cvi16008@cb.students.amrita.edu

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_165

detection system helps to keep driver alert, by showing the forthcoming road. Such a system would help to support drivers in recognizing any unsafe situations prior, to avoid road accidents, by sensing and understanding of the environment around and also by reducing the risk of increasing the speed of the vehicle.

Different techniques have been proposed to detect roads using image processing techniques. Majority of the methods use the edge detection done by using Canny filter, and finally the road boundary detection by Hough transform. This paper analyzes and compares two different methods for road detection. First method uses the widely used "Canny" edge detection technique with "Hough transform" for road detection. The second method is a slight variation of the first method. Instead of "Canny" edge detection, this algorithm uses "boundary extraction" technique for detection of road boundaries along with "Hough transform". This paper contains five sections; Sect. 1 is the introduction, and Sect. 2 talks about the related work done in this area. Section 3 explains the two proposed algorithms and their outputs in detail. Section 4 gives the experimental results of the system. The conclusions and future works are in Sect. 5.

2 Related Work

Kumar and Kaur [1] deal with a comprehensive review of different works in lane detection techniques. First, an image of road is taken with the help of a camera and converted to a grayscale image; then, the filters are applied to remove noise present in the image and the edge detection is done by using Canny filter. Finally, the lane boundary is detected by using Hough transform. Rasmussen [2] explains the unsupervised algorithm for ill-structured roads. Here, texture orientation is computed using "Gabor wavelet" filters. This technique is used to estimate straight roads and also for curved and undulating roads by vanishing point detection. Kong et al. [3] implement two main steps. First is the estimation of vanishing point of straight roads and curved roads and second is the constrained road segmentation, done using the detected vanishing point. The proposed approach adapts a Gabor filter-based adaptive soft voting scheme. Then, edge detection technique is used to detect road boundaries. The technique used here is AdaBoost-based region segmentation and boundary detection. Aly [4] proposed the lane detection process using Hough transform that detects edges to identify different lines separated over an image. Saha et al. [5] describe an automated approach of the road lane detection process. "Labeling" and "flood fill" algorithms are used for lane detection. Hu et al. [6] propose a system to check for cracks located at the surface of roads. Danti et al. [7] made an attempt to invent an automated driver guidance mechanism to make driving safe and easier on Indian roads. The proposed method is based on detecting the issues of Indian roads like potholes and then recognizes road signs in Indian roads. Xue et al. [8] proposed a real-time positioning method for robotic cars in urban areas using an efficient lane marking algorithm. An efficient shape registration algorithm is proposed to detect the distance between the detected lane markings which improves the accuracy of global localization of robotic car. Moeves and Kruse [9] proposed the usefulness of two recent trends in fuzzy methods of machine learning. First, fuzzy support vector machine is identical to a special type of SVM. Then, they categorize the existing approaches to develop fuzzy rules from SVM. Guan et al. [10] proposed a system for detecting road lanes in automotive applications. Here, they have explained an FPGA hardware implementation of novel Hough transform (HT) architecture. Aminuddin et al. [11] proposed videos footages that are considered for road detection dynamically. From the point of intersection of the road lane, the vanishing point is found out, and then, road triangles are formed. Then, using image multiplication technique between the original image and triangle-shaped mask image, road is detected by eliminating the area outside of the road boundary. Fernandes and Oliveira [12] claim Hough transform as a popular tool for detection of missing data. The proposed method presents an improved voting scheme for Hough transform in large images to achieve real-time performance on large images. This method improves the performance of the voting scheme significantly and creates a voting map which makes the transform more robust for detecting the lines. Hari et al. [13] proposed a new method called "midpoint Hough transform" which calculates the midpoints of the two selected random points of the input image which is in binary format. İf the midpoint is a bright point of the same image, then the point is transformed to the parameter space by using standard Hough transform equation. This process is repeated until no bright points are left in the image.

3 Proposed Algorithm

In this paper, mainly comparison and analysis of two different algorithms for road detection are being done. The first algorithm deals with the detection of roads using "Canny" edge detection and then applying "Hough" transform for detection of the boundary lines of the road. The second algorithm deals with the detection of straight roads by applying "boundary extraction" first to the input image and then applying the "Hough" transform for detecting the straight lines of the boundary of the roads. Here, the input dataset consists of 200 jpeg images of "straight urban" roads with minimum vehicles or obstacles. The dataset images are taken from a camera mounted inside front of a car. Both the algorithms are implemented using the "Matlab 2016b" tool.

3.1 Algorithm I: Straight Road Detection Using "Canny" Edge Detection and "Hough" Transform

The algorithm works in the following five steps:

Step 1: 3D Gaussian Filtering: The input image is first smoothed using 3D Gaussian filtering for the removal of existing noise present in the image. The presence of noise will lead to inappropriate detection of edges. The 3D Gaussian filter filters 3-D input

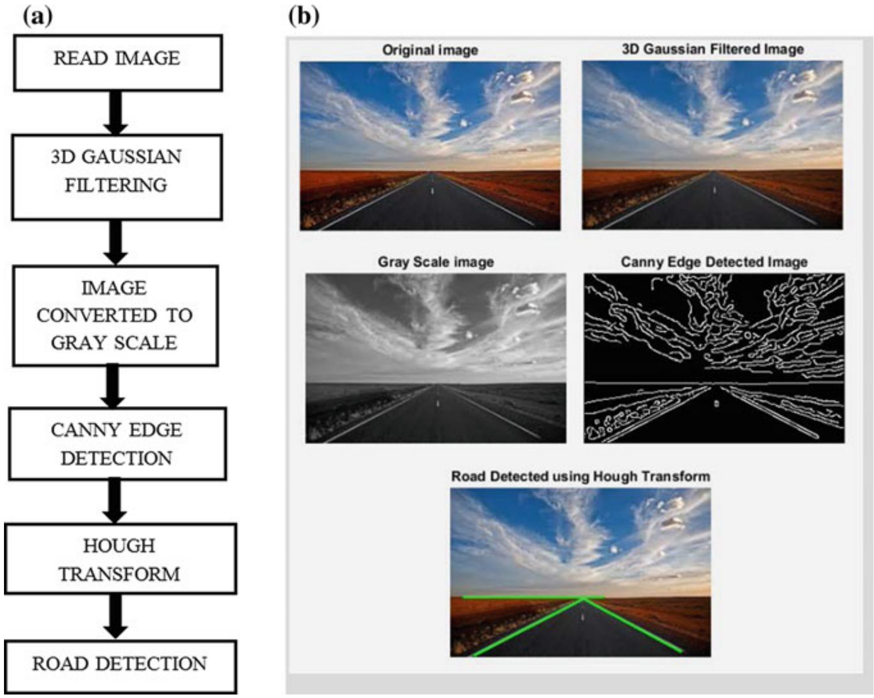


Fig. 1 **a** Architecture diagram for algorithm I; **b** output at each step

image with a 3-D Gaussian smoothing kernel. The filter is applied with a standard deviation of 0.5.

Step 2: Conversion to Grayscale: The image is then converted to grayscale image. This is done so that the time taken to process the image decreases.

Step 3: Canny Edge Detection: Then “Canny edge detection” technique is applied to the image to detect the edges present in the image. The Canny edge detector algorithm is a highly efficient algorithm that detects a large range of edges in a given image. It suppresses all the edges that are weak and outputs only the strong edges.

Step 4: Applying Hough Transform: To the Canny edge detected image, Hough transform is applied to detect the straight lines present in the image. Hough transform is a technique to extract features of objects that fall in a certain shape class by a procedure of voting.

Step 5: Road Detection: The Hough transform detects the lines that are present in the image. Here, in this algorithm, it detects the lines along the road boundaries and outputs an image with the detection of road in the input image.

The working of this algorithm can be depicted using the architectural diagram in Fig. 1a. The output at each step of the algorithm is given in Fig. 1b.

3.2 Algorithm II: Straight Road Detection Using "Boundary Extraction" and "Hough" Transform

The algorithm works in the following six steps:

Step 1: 3D Gaussian Filtering: As done in algorithm I, first the image is filtered or the noise is removed using 3D Gaussian filtering.
Step 2: Extracting the "Blue" Channel of the RGB Image: The filtered image is then split into its three channels: red channel, blue channel, and green channel. Our aim here is to detect roads, whose majority of the color information lies in the blue channel of the image. Hence, here we discard the red and the green channels. Further processing is done only on the "blue" channel of the image.
Step 3: Threshold the Blue Channel to Create a Blue Mask Image: The extracted blue band of the image is then further subjected to threshold to form the threshold image which we will call the "blue mask".
Step 4: Boundary Extraction: The "boundary extraction" algorithm is applied to this "blue mask" image to extract the boundaries of different regions present in the image.
Step 5: Applying Hough Transform: Finally, "Hough transform" for straight line detection is applied to the boundary extracted image to detect the different lines present in the image.
Step 6: Detection of Road Boundaries: The "Hough transform" algorithm detects the lines along the road boundaries and outputs an image with the detection of road in the input image.

The architecture diagram is shown in Fig. 2a. The intermediate results of the algorithm are shown in Fig. 2b.

4 Experimental Results and Analysis

Both the algorithms explained here were implemented and compared parallelly. A dataset of 200 different road images was given as input to each of the algorithms. The input image dataset did not contain any images of curved roads or roads with shadows. Road detection for each image was done using both the algorithms and the Hough transform lines formed for the road boundary detection, which were compared for each image of the dataset. The resultant "road-detected" images were manually analyzed and compared. Only those images where both the boundaries of the road were detected were considered as correct detection of the road. All the other images, where even if one of the road boundaries was detected, were not considered as correct road detection. Out of the total number of lines detected by Hough transform, only three lines were displayed on the image for road detection. Only those images which had all the three line aligned properly over the road boundaries were considered as correct detection of the road. All the other images were considered as wrong

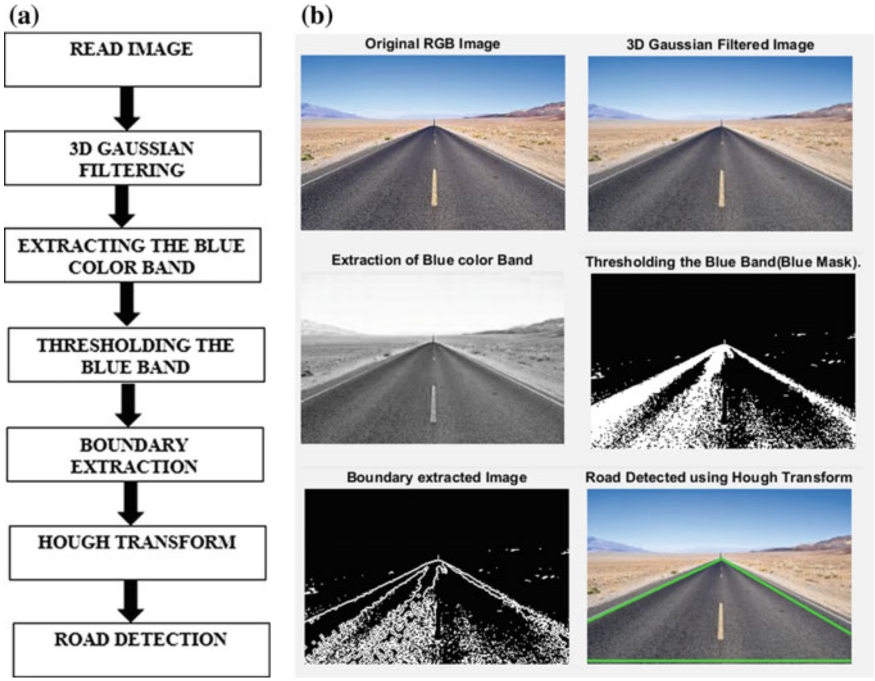


Fig. 2 **a** Architecture diagram of algorithm II; **b** output at each step

detection of road boundary. The algorithms were implemented using Matlab 2016b. The output of each algorithm for five different images is shown in Fig. 3.

The experiment results are shown in Table 1:

Table 1 shows the working of the two algorithms for 200 different road images. The dataset consists of images that are of straight roads taken from a camera mounted inside front of a car. The images are first subjected to Gaussian filtering for noise removal as preprocessing technique. The image dataset did not contain any images of any images of curved roads or roads with shadows.

Among the input images, those images where both the boundaries of the road were detected were considered as correct detection of the road. Out of the total number of lines detected by Hough transform, only three lines were displayed on the image for road detection. Only those images which had all the three lines aligned properly over the road boundaries were considered as correct detection of the road. All the other images, where even if one of the road boundaries was detected, were considered as wrong detection of road boundary.

As we can see in Table 1, the algorithm I which uses Canny edge detection along with Hough transform correctly detects roads for 151 images out of 200. In the remaining 49 images, the road was not detected properly or only a single boundary of the road was detected. In the second algorithm using Boundary extraction instead of Canny edge detection and Hough transform, road was correctly detected in 173

Table 1 Working of the algorithms for total of 200 images

Road detected correctly—Alg-I	Road detected wrongly—Alg-I	Accuracy = (correct detection/total images) * 100—for Alg-I (%)	Road detected correctly—Alg-II	Road detected wrongly—Alg-II	Accuracy = (correct detection/total images) * 100—for Alg-II (%)
151	49	75.5	173	27	86.5

Fig. 3 Output of both algorithms for five different images

images out of 200 images. In 27 images, the road was not detected properly by algorithm II.

The accuracy of both the algorithms was calculated by the following formula:

$$\text{Accuracy} = (\text{correct detection}/\text{total images}) * 100 \quad (1)$$

Accuracy got for algorithm I is 75.5% which is less as compared to the accuracy of 86.5% for algorithm II. This shows that the algorithm II works better for road detection, as compared to the algorithm I.

5 Conclusion

The extraction of straight roads from an image is very important requirement for preprocessing of images captured by surveillance cameras mounted within a vehicle. Existing road extraction methods widely make use of "Canny" edge detection technique for extraction of the edges in the roads. A new algorithm using "boundary extraction" with "Hough transform" for detection of boundaries of the roads is implemented in this paper. Road extraction using "Canny edge detector" yielded less accuracy than the "boundary extraction" based road detection. Thus, for straight urban roads, it is found that boundary extraction-based Hough transform produced better results.

Future work would include the extension of this algorithm for road images having shadows and for curved and muddy roads.

References

1. Kumar D, Kaur G (2008) Lane detection techniques: a review. IEEE Trans Intell Transp Syst 9(1):16–26
2. Rasmussen C (2004) Texture-based vanishing point voting for road shape estimation. In: BMVC, pp 1–10
3. Kong H, Audibert JY, PonceJ (2009) Vanishing point detection for road detection. In: IEEE conference on computer vision and pattern recognition, 2009. CVPR 2009. IEEE, pp 96–103
4. Aly M (2008) Real time detection of lane markers in urban streets. In: Intelligent vehicles symposium, 2008 IEEE, pp 7–12
5. Saha A, Roy DD, Alam T, Deb K (2012) Automated road lane detection for intelligent vehicles. Glob J Computer Sci Technol
6. Hu H, Gu Q, Zhou J (2010) HTF: a novel feature for general crack detection. In: 2010 17th IEEE international conference on image processing (ICIP). IEEE, pp 1633–1636
7. Danti A, Kulkarni JY, Hiremath PS (2012) An image processing approach to detect lanes, pot holes and recognize road signs in Indian roads. Int J Model Optim 2(6):658
8. Cui D, Xue J, Zheng N (2016) Real-time global localization of robotic cars in lane level via lane marking detection and shape registration. IEEE Trans Intell Transp Syst 17(4):1039–1050
9. Moewes C, Kruse R (2011) On the usefulness of fuzzy SVMs and the extraction of fuzzy rules from SVMs. In: EUSFLAT Conf., pp 943–948

10. Guan J, An F, Zhang X, Chen L, Mattausch HJ (2017) Parallelization of Hough transform for high-speed straight-line detection in XGA-size videos. In: 2017 IEEE international conference on consumer electronics-Taiwan (ICCE-TW). IEEE, pp 313–314
11. Aminuddin NS, Masrullizam MI, Ali NM, Radzi SA, Saad WHM, Darsono AM (2017) A new approach to highway lane detection by using hough transform technique. J Inf Commun Technol 16(2):244
12. Fernandes LAF, Oliveira MM (2008) Real-time line detection through an improved Hough transform voting scheme. Pattern Recogn 41(1):299–314
13. Hari CV et al (2009) Mid-point hough transform: a fast line detection method. In: India conference (INDICON), 2009 Annual IEEE

Investigating the Impact of Various Feature Selection Techniques on the Attributes Used in the Diagnosis of Alzheimer's Disease

S. R. Bhagyashree and Muralikrishna

Abstract According to the Dementia India report 2010, it is estimated that over 3.7 million people are affected by dementia and is expected to be double by 2030. Around 60–80% of the demented are suffering from Alzheimer's disease. Neuropsychological tests are useful tools for diagnosis of dementia. Diagnosis of dementia using machine learning for low- and middle-income setting is a rare study. Various attributes are used for diagnosing dementia. Finding the prominent attributes among them is a tedious job. Chi-squared, gain ratio, info gain and ReliefF filtering techniques are used for finding the prominent attributes. Cognitive score is identified as the most prominent attribute.

1 Introduction

Dementia is a neurodegenerative disease. It causes loss of cognitive functions such as reasoning, memory and other mental abilities which may be due to trauma or normal ageing [1]. It is estimated that 10–20% of people aged above 65 are having mild cognitive impairment [2]. It has been estimated that worldwide, around 4.6 million new dementia cases are identified every year [3]. Approximately, 473,000 people with age 65 and above will develop Alzheimer's disease [4]. Neuropsychological assessment is used for diagnosis of the disease. Alzheimer Disease International has designed a battery named Community Screening Instrument for Dementia (CSID) for diagnosis of dementia [5]. For the same dataset, classification using Naïve Bayes is performed and it has fetched a classification accuracy of 96.69%. Wrapper method is used as feature selection technique. The Cog-Score is identified as the prominent

S. R. Bhagyashree (✉)
E & C Department, ATME College of Engineering, Mysore, Karnataka, India
e-mail: srbhagyashree@yahoo.co.in

Muralikrishna
Wellcome DBT Allianz,
CSI Holdsworth Memorial Hospital, Mysore, India
e-mail: muralidoc@gmail.com

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_166

attribute [6]. In this work, the feature selection is performed using filter method. In this work, the impact of various feature selection techniques on the accuracy of classification before and after feature selection techniques is studied. Along with that, the impact of number of attributes on classification accuracy is also studied. This work also focuses on the most prominent attribute that is needed to diagnose dementia. The paper is organized as follows. Section 2 explains the architecture of the proposed work which includes details of dataset, preprocessing with related literature survey and classification. Section 3 has results and discussion. Section 4 focuses on conclusion.

2 Related Work

Dementia is classified into Alzheimer's disease, dementia with Lewy bodies, Parkinson's disease, normal pressure hydrocephalus, cascular dementia and front temporal labour degeneration dementia. Alzheimer's disease is the major stakeholder [7]. The disease can be diagnosed by consulting general physician, by conducting neuropsychological assessment and by doing magnetic resonance imaging. Common brief assessments include the Mini-Mental State Examination (MMSE), the Brief Cognitive Rating Scale and Alzheimer's Disease Assessment Scale-Cognitive (ADAS-Cog), mini-cog, Montreal Cognitive Assessment (MOCA) etc. These neuropsychological assessments have their own disadvantages [8]. Alzheimer's Disease International (ADI) has founded 10/66 research group with an objective to design a battery which will overcome all the above problems. The group has studied the subjects of various age groups in different developing countries and designed a battery called Community Screening Instrument for Dementia (CSID) [5]. In this work, CSID battery is used for diagnosis.

Availability of specialist human resource including psychologists is less than one for one lakh population, whereas the demented is around 4 million and it is increasing with time [9]. This infers that at no point of time all the demented get diagnosed on time. Hence, there is a need of method which covers maximum population and diagnoses them on time.

Progress in data mining applications and its implications are manifested in the areas of information management in healthcare organizations, health informatics, epidemiology, patient care and monitoring systems, assistive technology, large-scale image analysis to information extraction and automatic identification of unknown classes. According to the most recent and sound systematic literature overview, performed by Esfandiari et al., four main application areas of DM application in medicine can be defined [10].

Increasing the efficiency and elimination of the human factor which deals with tasks for diagnosis of certain diseases where accuracy is essential is one such application. Researchers have used machine learning algorithms for diagnosis of various diseases like liver disorder, heart diseases etc. [11, 12].

3 Proposed Work

The architecture of the work is shown in Fig. 1. The dataset comprises details of 466 subjects having 50 attributes. In preprocessing, the feature selection is done using various filtering techniques. The classification is done using Naïve Bayes. The model evaluation is performed using 10-fold cross-validation.

3.1 Collection of Dataset

Global cognitive function is measured by administering the Community Screening Instrument for Dementia (CSID) which includes a 32-item cognitive test comprises 50 attributes. This includes assessing, orientation, comprehension, memory, naming, language expression, etc. Dataset consists of details of 282 men and 184 women, aged between 60 and 90 with different levels of education starting from illiterates to postgraduates, belonging to different religions.

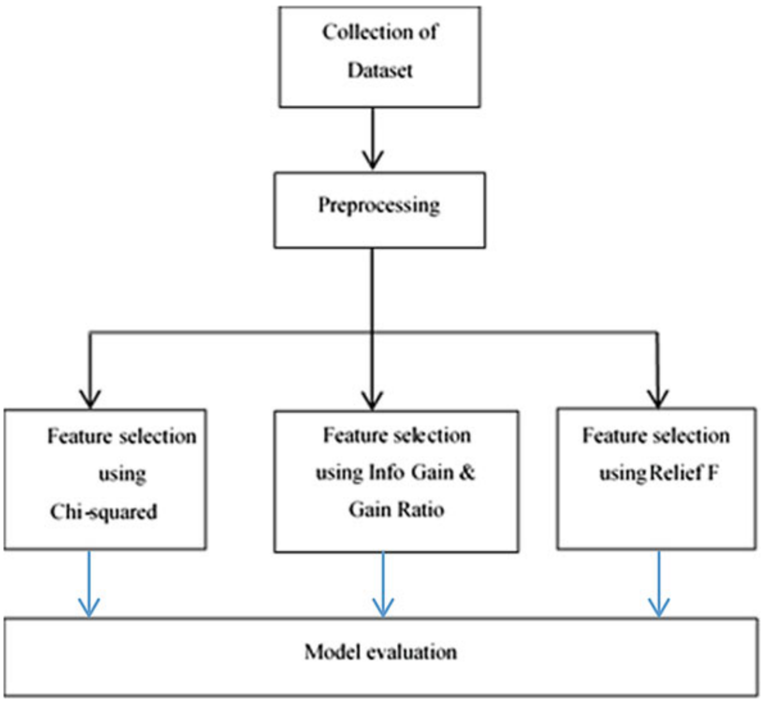


Fig. 1 Architecture

3.2 Preprocessing

Preprocessing is very important step in discovering knowledge from data. The initial steps like data cleaning and data integration are not performed as the data is a manually collected and is free from missing of any information. The data reduction is done using feature selection. The filter approach is used for implementing the same.

1. Attribute evaluation algorithms
2. Subset evaluation algorithms.

The algorithms are categorized based on whether they rate the relevance of individual features or feature subsets. Attribute evaluation algorithms rank the features individually and assign a weight to each feature according to each feature's degree of relevance to the target feature. The attribute evaluation methods are likely to yield subsets with redundant features since these methods do not measure the correlation between features. Subset evaluation methods, in contrast, select feature subsets and rank them based on certain evaluation criteria and hence are more efficient in removing redundant features [13, 14].

3.2.1 Feature Selection Using Filter Method

The filter attribute selection method is independent of the classification algorithm. Filter method is further categorized into the following techniques.

3.2.2 Information Gain Attribute Evaluation

The entropy and the information gain are attribute measures which indicate how much percentage the given attribute separates the training dataset according to their final classification. The entropy and gain for a set S are calculated as

$$\text{Entropy}(S) = \sum_{i=1}^{n} -P_i \log_2 P_i \tag{1}$$

$$\text{Gain}(A) = \text{Entropy}(S) - \sum_{k=1}^{m} \frac{|S_k|}{|S|} \times \text{Entropy}(S_k) \tag{2}$$

where

"n" is the number of class.
P_i is the probability of S belongs to class i.
S_k is the subset of S [8].

3.2.3 Gain Ratio (GR) Attributes Evaluation

The gain ratio is the non-symmetrical measure that is introduced to compensate for the bias of the *IG* [15]. *GR* is given by

$$GR = \frac{IG}{H(X)} \tag{3}$$

As Eq. (3) presents, when the variable *Y* has to be predicted, we normalize the *IG* by dividing the entropy of *X,* and vice versa. Due to this normalization, the *GR* values always fall in the range [0, 1]. A value of $GR = 1$ indicates that the knowledge of *X* completely predicts *Y* and $GR = 0$ means that there is no relation between *Y* and X. In opposition to *IR*, the *GR* favours variables with fewer values [16].

3.2.4 Chi-Squared Statistic

This method measures the lack of independence between a term and the category. Chi-squared is the common statistical test that measures divergence from the distribution expected, if one assumes that the feature occurrence is actually independent of the class value. As a statistical test, it is known to behave erratically for very small expected counts, which are common in text classification. It is because of having rarely occurring word features and having few positive training examples for a concept. In statistics, the X^2 test is applied to test the independence of two events, where two events A and B are defined to be independent if $P(AB) = P(A)P(B)$ or, equivalently, $P(A|B) = P(A)$ and $P(B|A) = P(B)$. In feature selection, the two events are occurrence of the term and occurrence of the class. Feature selection using the X^2 statistic is analogous to performing a hypothesis test on the distribution of the class as it relates to the values of the feature in question [17].

3.2.5 ReliefF

ReliefF attribute evaluation evaluates the worth of a feature by repeatedly sampling an instance and considering the value of the given feature for the nearest instance of the same and different classes. This attribute evaluation assigns a weight to each feature, based on the ability of the feature to distinguish among the classes and then selects those features whose weights exceed a user-defined threshold as relevant features. Chi-squared, entropy and CFS filters are used for feature selection for two different datasets. The first group comprises 327 samples of acute lymphoblastic leukaemia, and the second group consists of 162 samples of ovarian cancer. The dataset is classified using classification algorithms, namely, k-NN, C4.5 and Naïve Bayes SVM [18].

3.3 Classification

Features are selected using different filtering techniques. Naïve Bayes and C4.5 are part of top ten algorithms used in data mining [19]. In the current study, the data is classified using Naïve Bayes algorithm. Naive Bayesian classifier is a selective classifier which calculates the set of probabilities by counting the frequency and combination of values in a given dataset. It assumes that all variables which contribute towards classification are mutually independent [20].

4 Results and Discussion

The details of 466 subjects are collected by conducting neuropsychological test. The data so collected is realistic, manually collected and hence the data is free from noise. The data is converted to ARFF format. The feature selection is done by using filter approach. Chi-squared, gain ratio, info gain and ReliefF are used as attribute evaluators. The dataset is classified using Naïve Bayes. The model evaluation is done by applying 10-fold cross-validation.

Table 1 shows the classification accuracy after applying Naïve Bayes algorithm for datasets using different feature selection filters. The number of attributes considered is 50, 35, 25, 15 and 8, respectively.

Table 2 shows the results before and after feature selection.

Table 1 Results after applying Naïve Bayes classification

Classification accuracy (%)					
Filters	Number of attributes				
	50	35	25	15	8
Chi-squared	96.78	96.35	96.56	96.56	98.06
Gain ratio	96.78	96.35	97.21	96.78	98.08
Info gain	96.78	96.35	96.99	97.42	96.99
ReliefF	96.78	96.13	97.85	96.99	94.63

Table 2 Classification Accuracy before and after feature selection

Filters	Classification accuracy (%) Naïve Bayes	
	Before feature selection	After feature selection
Chi-squared	96.78	98.06
Gain ratio	96.78	98.06
Info gain	96.78	96.99
ReliefF	96.78	94.63

Table 3 Results before and after feature selection

Rank	Chi-squared	Gain ratio	Info gain	ReliefF
1	Cog-Score	Cog-Score	Cog-Score	Sentence repeat
2	Orientation with time	City	Orientation with time	Year
3	Story	Pray	Story	Long mem
4	Month	Watch	Semantic memory	Month
5	Semantic memory	Chemis	Month	Elbow
6	WLR	Learn2	Year	Nrecal
7	Learn2	Orientation with time	WLR	Street

Table 3 shows the list of high ranking attributes that are selected with the application of different feature selection techniques. Top seven attributes are listed here with Class as eighth attribute.

5 Conclusion

In this work, the details of 466 subjects comprising orientation, comprehension, memory, naming and language expression are considered. In preprocessing, the feature selection is performed using filter method. In this work, chi-squared, gain ratio, info gain and ReliefF are used as attribute evaluators. The features were selected using ranker approach. The attributes with lower ranking were removed and the classification accuracy is calculated at every stage.

The Naïve Bayes classifier shows a variation in the classification accuracy with variation in number of attributes. In chi-squared, the attributes 25 and 15 act as threshold values as there is change in a classification accuracy below and above these values. In chi-squared and gain ratio, maximum classification accuracy is achieved with minimum number of attributes. When the number of attributes is decreased from 50 to 35, the classification accuracy has decreased in all the evaluator schemes. Further, with an additional decrease of 10 attributes, there is an increase in the accuracy of the classification. The classification accuracy is more for less number of attributes and vice versa except for ReliefF. This infers that to achieve higher efficiency, the number of attributes needs to be minimal.

Table 3 clearly shows the significance of Cog-Score and orientation with time as prominent attributes in most of the feature selection techniques. The Cog-Score represents the entire set of attributes; hence, it is considered as the vital attribute for the accurate diagnosis of dementia. By considering the previous work [6] and the present work, it is very much clear that, irrespective of the feature selection techniques used, Cog-Score acts as the prominent attribute in the diagnosis of Alzheimer's disease.

Acknowledgements We are grateful to the participants and their family members for taking part in this study. Our sincere thanks to Principal and Management of ATME College of Engineering.

Ethics The ethics approval for this study was obtained from the Ethics Committee of CSI Holdsworth Memorial Hospital, Mysore.

Only those participants who were able to provide fully informed consent participated in this study and the informed consent was obtained from the participants.

Conflict of Interest None of the authors have any conflict of interest to declare.

References

1. Bhagyashree SR, Sheshadri HS (2014) An initial investigation in the diagnosis of Alzheimer's disease using various classification techniques. IEEE international conference on computational intelligence and computing research (ICCIC), pp 1–5
2. Gaugler J, James B, Johnson T, Scholz K, Weuve J, (2015) Sc.D Alzheimer's Association 2015 Alzheimer's disease facts and figures. Alzheimer's & Dementia 2015, pp 1–83
3. Prince M, Bryce R, Albanese E, Wimo A, Ribeiro W, Ferri CP (2013) The global prevalence of dementia: a systematic review and metaanalysis. Alzheimer's dement 9(1):63–75.e2. https://doi.org/10.1016/j.jalz.012.11.007
4. Bhagya Shree SR, Sheshadri HS (2014) An approach to preprocess data in the diagnosis of Alzheimer's disease. IEEE international conference on cloud computing and internet of things (CCIOT 2014), pp 135–139
5. Luisa AL et al (2009) Population normative data for the 10/66 Dementia Research Group cognitive test battery from Latin America, India and China: across sectional survey. Access NIH public, PubMed central, BMC Neurol 9:1–11
6. Bhagya Shree SR, Sheshadri HS (2016) Diagnosis of Alzheimer's disease using Naive Bayesian Classifier. NCAA, Neural Comput Appl 29:1–10
7. Bhagya Shree SR, Sheshadri HS, Joshi S (2014) A review on the method of diagnosing Alzheimer's disease using data mining. Int J Eng Res Technol 3(3):2417–2420. ISSN: 2278–0181
8. Shree SB, Sheshadri H (2014) An initial investigation in the diagnosis of Alzheimer's disease using various classification techniques. In: 2014 IEEE international conference on computational intelligence and computing research (ICCIC). IEEE, pp 1–5
9. Shaji K, Jotheeswaran A, Girish N, Bharath S, Dias A, Pattabiraman M, Varghese M (2010) Alzheimer's and related disorders society of India. The Dementia India Report: prevalence, impact, costs and services for Dementia
10. Esfandiari N, Babavalian MR, Moghadam A-ME, Tabar VK (2014) Knowledge discovery in medicine: current issue and future trend. Expert Syst Appl 41(9):4434–4463
11. Taneja A et al Heart disease prediction system using data mining techniques. Orient J Comput Sci Technol 6(4):457–466
12. Rajeswari P, Reena GS (2010) Analysis of liver disorder using data mining algorithm. Global J Comput Sci Technol 10(14):48
13. Henderson AS, Jorm AF (2000) Definition of epidemiology of dementia: a review. Dementia 2:1–33
14. Vanaja S, Kumar KR (2014) Analysis of feature selection algorithms on classification: a survey. Int J Comput Appl 96(17):975

15. Hussain L, Aziz W, Nadeem S, Abbasi A (2014) Classification of normal and pathological heart signal variability using machine learning techniques
16. Novaković J, Štrbac P, Bulatović D (2011) Toward optimal feature selection using ranking methods and classification algorithms. Yugoslav J Oper Res 21(1). ISSN: 0354-0243 EISSN: 2334-6043
17. Ladha L, Deepa T (2011) Feature selection methods and algorithms. Int J Comput Sci Eng 1(3):1787–1797
18. Yu J, Hu S, Wang J, Wong GKS, Li S, Liu B, Deng Y, Dai L, Zhou Y (2002) A draft sequence of the rice genome (Oryza sativa L. ssp. indica). American Association for the Advancement of Science
19. Wu X, Kumar JV, Quinlan R, Ghosh J, Yang Q, Motoda H, Geoffrey J, Ng MN, Liu B, Yu PS, Zhou Z-H, Steinbach M, Hand DJ, Steinberg D (2008) Top 10 algorithms in data mining. Springer, pp 1–37
20. Ferreira D et al (2012) Applying data mining techniques to improve diagnosis in neonatal jaundice. BMC Med Inform Decis Making 1–5

Detection of Sleep Apnea Based on HRV Analysis of ECG Signal

A. J. Heima, S. Arun Karthick and L. Suganthi

Abstract Sleep apnea is a breathing disorder which occurs during sleep. Sleep apnea causes more health-threatening problems such as daytime sleepiness, fatigue and cognitive problems, coronary arterial disease, arrhythmias, and stroke. However, there is an extremely low public consciousness about this disease. The most common type of sleep apnea is *obstructive sleep apnea (OSA)*. Polysomnography (PSG) is the widely used technique to detect OSA. *Obstructive sleep apnea* is extremely undiagnosed due to the inconvenient and costly polysomnography (PSG) testing procedure at hospitals. Moreover, a human expert has to monitor the patient overnight. Hence, there is a requirement of new method to diagnose sleep apnea with efficient algorithms using noninvasive peripheral signal. This work is basically aimed at detection of sleep apnea using a physiological signal electrocardiogram (ECG) alone which is taken from free online apnea ECG database provided by PhysioNet/PhysioBank. This database consists of 70 ECG recordings. A detailed time- and frequency-domain features and nonlinear features extracted from the RR interval of the ECG signals for observing minutes of sleep apnea are occurred in this work. Time-domain features mean HR ($P = 0.0093$, $r = 0.3593$) and RR interval mean (ms) ($p = 0.0003$, $r = 0.376$), frequency-domain features VLF power (%) ($P = 0.00659$, $r = 0.1081$) and HF power (%) ($P = 0.00135$, $r = 0.41138$), and nonlinear analysis feature SD1 ($P = 0.00039$, $r = 0.18998$), significantly different for normal and apnea ECG. Further supervised learning algorithms have been used to classify ECG signal to differentiate normal and apnea data. The overall efficiency is 90.5%. Algorithms which deal ECG signal along with respiratory signal will give more incite about apnea disease and also the classification accuracy may be improved.

A. J. Heima · S. Arun Karthick (✉) · L. Suganthi
Department of Biomedical Engineering, SSN College of Engineering, Chennai, India
e-mail: arunkarthicks@ssn.edu.in

A. J. Heima
e-mail: heima16455004@bme.ssn.edu.in

L. Suganthi
e-mail: suganthil@ssn.edu.in

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_167

1 Introduction

Sleep constitutes a naturally send back to the state of rest for the body and mind. The central nervous system is restored during sleeping period. Therefore, sleep plays an important role in good physical condition by defending the mental and physical health. It also increases the quality of life. Human sleeps for about one-third of their whole lifetime.

People with sleeping disorder could not sleep peacefully. Deficiency of sleep can affect the body physically, emotionally, and psychologically. Eighty-four kinds of sleep disorders have been found till date. The most common sleep disorders are sleeplessness, sleep apnea, narcolepsy, and restless leg syndrome. A common type of sleep disorder is constituted by sleep apnea. Sleep apnea is defined as a pause in breathing for 10 s or more during sleep. The three types of sleep apnea are obstructive sleep apnea (OSA), central sleep apnea (CSA), and mixed apnea. The restriction on airflow, which conducts snoring during sleep, causes OSA. The brain failed to send signal breathing demand to the muscle during sleep which causes CSA. Mixed apnea is a combination of OSA and CSA. Obstructive sleep apnea is the most common type of sleep apnea. The effects of OSA are tiredness during the day, reduced reaction times, impaired vision or cognitive problems, and memory disorders [1]. Moreover, some studies show an enlarged number of apnea episodes that can enlarge the risk of diabetes [1]. OSA is also linked to cardiovascular diseases. Apnea episodes may cause cardiac ischemia [1].

In general population, 14% men and 5% women have obstructive sleep apnea [2]. Most OSA cases are undiagnosed and cause 980 deaths each year, because of the inconvenience, expenses, and unavailability of testing. Polysomnography (PSG) is the usual testing process, which is a standard procedure for all sleep disorder diagnosis. It records the breath airflow, respiratory movement, oxygen saturation, body position, electroencephalogram (EEG), electrooculogram (EOG), electromyogram (EMG), and electrocardiogram (ECG). The dependency on PSG needs to be taken away from the home for effective detection and treatment of sleep apnea. Therefore, it is essential to develop simple and efficient methods which diagnose OSA [3].

ECG signals have been used in earlier research works for the detection of sleep apnea [4–7]. Heart rate variability in a huge sample of healthy subjects was analyzed using time- and frequency-domain methods and a Poincaré plot. In the Fourier transform of ECG-RR interval, power of ultralow frequency (ULF), power of very low frequency (VLF), and power of low frequency (LF) are good indicators of health in cardiac patients [8]. Mean, standard deviation, median interquartile, and mean absolute deviation are other features which are extracted from RR interval of ECG signal [9].

This work is mainly aimed at diagnosis of sleep apnea from physiological signal electrocardiogram (ECG) by using time- and frequency-domain parameters of ECG–RR interval and nonlinear parameter derived from the Poincaré plot of RR interval.

2 Materials and Methods

The vital features of ECG and OSA show a high interaction among all biosignals. In this research, we use the free online apnea–ECG database provided by PhysioNet/PhysioBank. It contains 70 overnight ECG recordings provided by Philipps University, Marburg, Germany. The ECG signal is acquired at 100 samples per second with an amplitude resolution of 5 μV per least significant bit. Each record contains 8 h of ECG signal, and a record of diagnoses of sleep apnea per minute by clinical experts. In the ECG recordings, each apnea group recording has 100 or more minutes with tangled breathing. Twenty-five recordings were selected in this category (24 male, 1 female) with mean and standard deviation of age 51.84 ± 6.19. In control group, recordings have fewer than 5 min of disordered breathing. Seventeen recordings were selected in this category (10 male, 7 female) with mean and standard deviation of age 33.12 ± 5.74. Figure 1 shows the block diagram of methodology.

The ECG signal was affected by different noises during its acquisition and transmission. So, ECG signal has to be preprocessed. Filtering techniques were used to remove these noises. The AC component of the ECG signal is taken out using an analog low-pass filter with a cutoff frequency of 20 Hz. Then, the DC component of the signal is removed using a high-pass filter with a cutoff frequency of 0.15 Hz. We have considered 1 min data (at least 60 successive cardiac cycles) in order to reduce the influence of the respiratory system and other movement artifacts, and then the parameters are extracted and averaged. An averaging period covering at least 60 heartbeats (1 min data) has been used in earlier studies to improve conviction in the RR interval parameter extracted from the ECG signal [9]. After the preprocessing of ECG signal, QRS detection algorithm has been implemented for the detection of RR interval. The following operations are carried out in QRS detection algorithm such as cancelation DC drift and normalization, low-pass filtering (cutoff frequency 35 Hz),

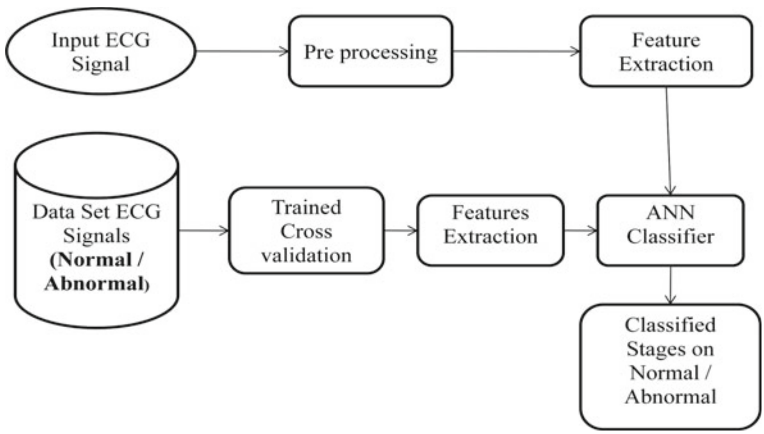


Fig. 1 Block diagram of methodology

high-pass filtering (cutoff frequency 5 Hz), derivative filter, squaring, and moving window integration [10]. In the first step of the algorithm, the signal is filtered by a low-pass and a high-pass filter in order to remove the influence of the muscle noise, the power line interference, the baseline wander, and the T-wave interference. Differentiation is performed to find the slope of the R-wave in the QRS complex. After finding the derivative, squaring operation is performed to find the characteristics of the signal. After the squaring of derivative of the signal, the start of the QRS complex is found with an adjusting threshold which is shown in Fig. 2. The peak of the squared signal is identified as the R-peak of the ECG data.

Kubios HRV has been used for performing time-domain, frequency-domain, and nonlinear analysis of RR intervals. The RR interval series is converted into equidistantly sampled series by interpolation (cubic spline interpolation) methods. There are many features which are extracted from ECG signal in time domain such as mean HR, SDNN, RR interval mean, and NN50.

In the frequency-domain analysis of RR interval, the power spectral density estimation is performed using fast Fourier transform (FFT) of RR interval. FFT spectrum is found based on Welch's periodogram method with 256 window size and 50% overlap. The frequency-domain features are very low-frequency (VLF) power, high-frequency (HF) power, and low-frequency (LF) power.

Poincaré plot is a nonlinear method to find correlation between successive RR intervals, (plot of RRj + 1 as a function of RRj). The parameters standard deviation 1 (SD1) and standard deviation 2 (SD2) are derived from Poincaré plots.

MATLAB pattern recognition tool is used for the classification of normal and apnea ECG. The pattern recognition tool uses supervised learning algorithms. The proposed algorithm classifies the signal on the basis of supervised learning algorithms. Extracted features from ECG signal are used for apnea classification.

Due to the non-uniformity of data, for nonparametric data analysis, Mann–Whitney U test is selected for the finding significant difference between the mean parameters of control and apnea group. We have calculated Spearman's rank correlation

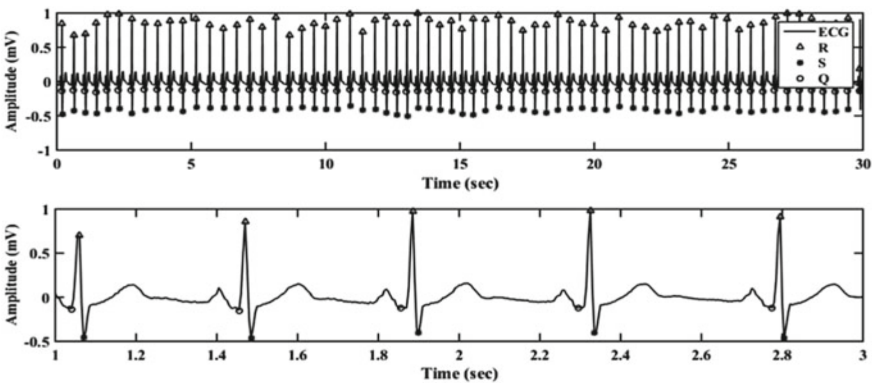


Fig. 2 ECG QRS detection

coefficient and probability using MATLAB 2007. The significant result was considered as $P > 0.05$.

3 Results

Using box plots, we have found the outliers for each parameter of normal and apnea ECG signals which are shown in Figs. 3, 4, and 5.

Without considering these outliers, the mean and standard error of the time-domain, frequency-domain, and nonlinear parameters are calculated which are shown in Figs. 6, 7, and 8, respectively. Among those features, mean HR ($p = 0.0093$, $r =$

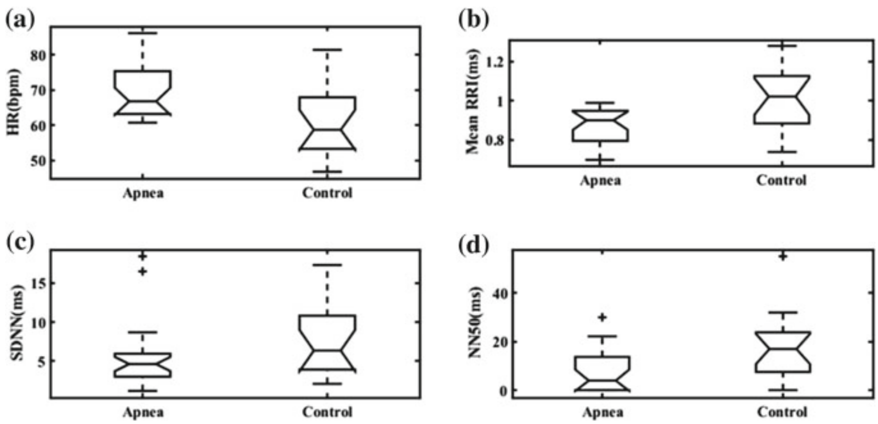


Fig. 3 Box plot for time-domain features, **a** heart rate (bpm), **b** mean RR interval (ms), **c** SDNN, **d** NN50

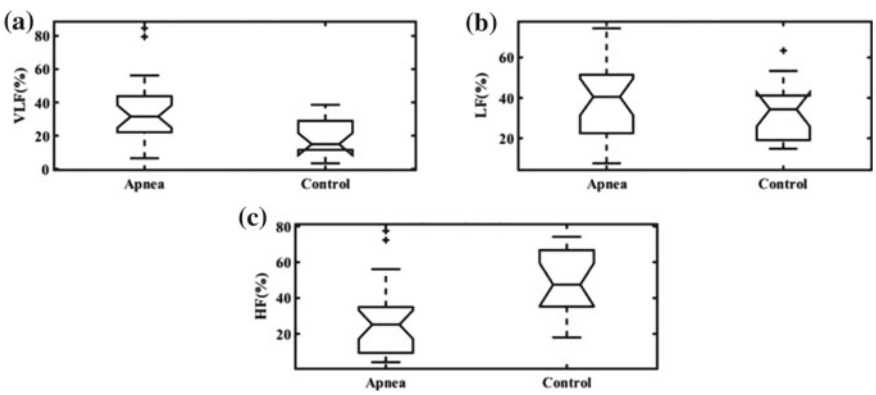


Fig. 4 Box plot for frequency-domain features, **a** very low-frequency (VLF) power, **b** high-frequency (HF) power, and **c** low-frequency (LF) power

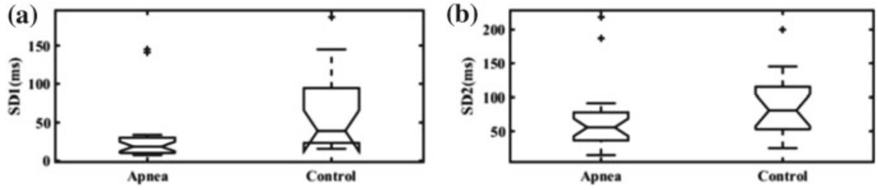


Fig. 5 Box plot for nonlinear features, **a** standard deviation 1 (SD1), **b** standard deviation 2 (SD2)

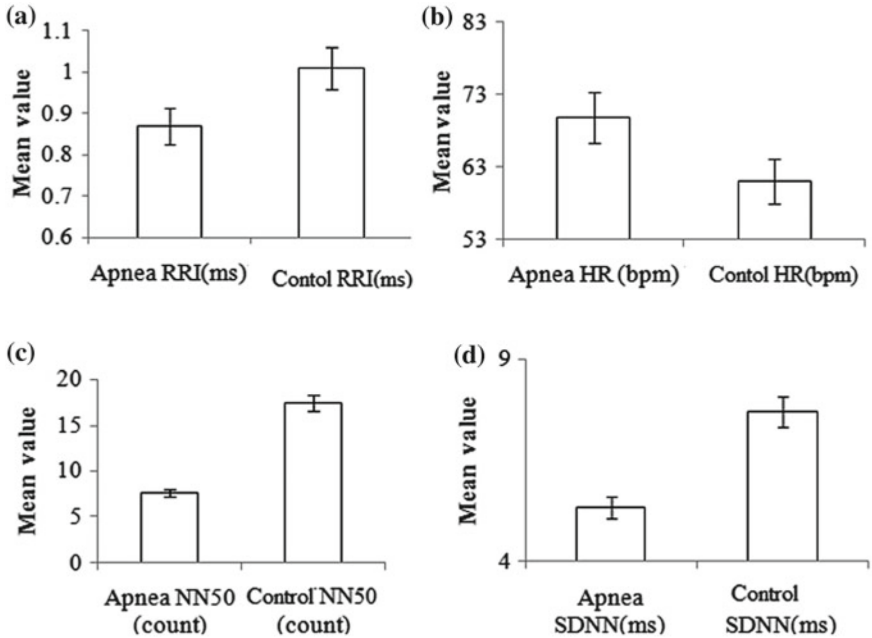


Fig. 6 Mean and standard error plot for time-domain feature. **a** RR interval, **b** Heart rate, **c** NN50, **d** SDNN

0.3593), RR interval mean (ms) (p =0.0003, r =0.376), VLF power (%) (p = 0.00659, r =0.1081), HF power (%) (p =0.00135, r =0.41138), and SD1 (P = 0.00039, r =0.18998) show significant difference between normal and apnea ECG.

Mean HR of normal person is lesser than apnea patient. HF of apnea person is lesser than normal patient. SD1 of apnea patient is lesser than normal person. The statistical analysis has been performed by using Mann–Whitney U test. The Spearman's rank correlation coefficient (r_c) and probability (P) which we obtained are given in Table 1. Artificial neural network algorithm which is written in MATLAB software classifies the significant features of test input. Classification is performed on the basis of supervised learning algorithms. Seventy percentage of input data is considered for the training input of ANN algorithm, and 30% of data is considered

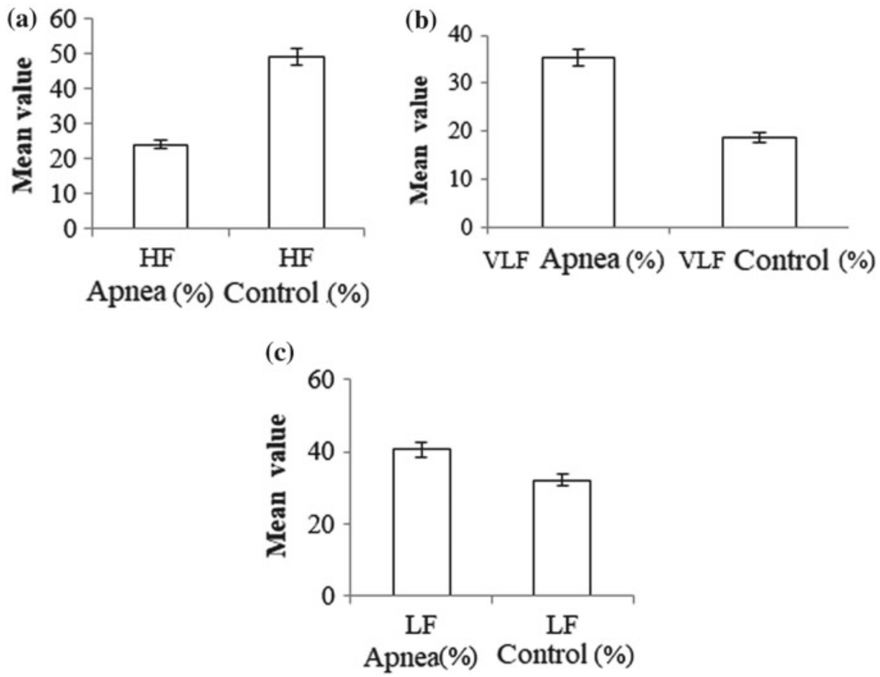


Fig. 7 Mean and standard error plots for frequency-domain features, **a** high-frequency (HF) power, **b** very low-frequency (VLF) power, and **c** low-frequency (LF) power

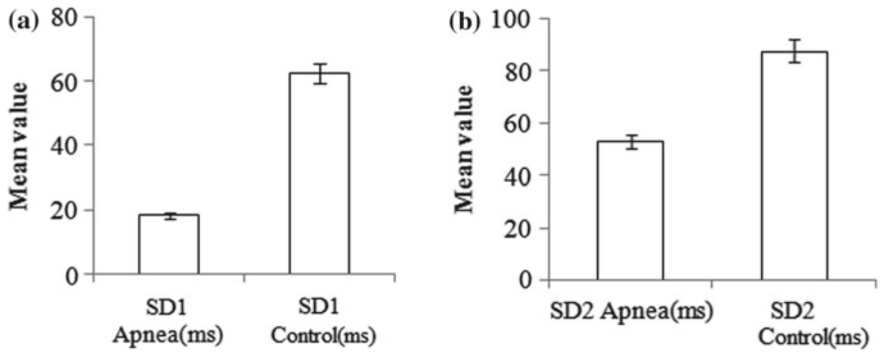


Fig. 8 Mean and standard error plot for nonlinear feature such as **a** standard deviation 1 (SD1), **b** standard deviation 2 (SD2)

for testing. The results are derived from the confusion matrix of normal and apnea patients as shown in Table 2. Performance measure of classification of normal and apnea signal is determined using accuracy, precision, sensitivity, and specificity. Percentage of each parameter describes the accuracy and efficiency for normal and apnea signal. The overall classification efficiency is 90.5% as shown in Table 2.

Table 1 Mean parameter values, Pearson correlation coefficient (r) and probability (P-value) using Mann–Whitney U test for control and apnea group

Feature	Normal	Apnea	Pearson correlation coefficient	P-value
Mean HR (bpm)	67.30116	85.84568	0.35931	0.0093*
SDNN (ms)	4.983797	4.615369	0.354	0.0646
RR interval mean(ms)	0.891515	0.698929	0.376	0.000*
NN50 (count)	13	6	0.2251	0.0646
Very low-frequency power (%)	23.7	10.1	0.1081	0.00659*
Low-frequency power (%)	69.4	45.1	−0.264	0.23846
High-frequency power (%)	7	44.9	0.41138	0.00135*
Standard deviation 1 (ms)	10.3	32.5	0.18998	0.00039*
Standard deviation 2 (ms)	34.1	62.9	0.08017	0.08606

*P-value < 0.05

Table 2 Classification accuracy of apnea and normal patient

	Prediction			Measurement
		Apnea	Control	
Actual	Apnea	True positive (TP) 24(57.1%)	False negative (FN) 3(7.1%)	Sensitivity 88.9%
	Control	False positive (FP) 1(2.4%)	True negative (TN) 14(33.3%)	Specificity 93.3%
Measurement		Positive predictive value 96%	Negative predictive value 82.4%	Accuracy 90.5%

4 Discussions

In apnea, respiratory rate varies with respect to the effect of disorder. The movement of diaphragm decides the position of the heart. Because the central tendon of the diaphragm is attached to the outer layer of heart (pericardium), the systemic or pulmonary arterial pressure or both increase during apnea. OSA patients have higher degree of pulmonary hypertension. Hemodynamic changes affect right and left ventricle during OSA episodes [11]. Because of these conditions, features derived from ECG signal vary during apnea [12].

The detailed study of HRV signal leads to better understanding of apnea. Recent studies show that an efficient algorithm that combines magnitude and frequency information from the power spectrum is more suitable for OSA detection [13]. The time-domain features related to OSA are proposed by de Chazal et al. [14]. Heart rate variability contains many features in both time and frequency domains and needs a different method in calculation of these features [15].

Changes in respiratory patterns can also vary HR and HRV independent of cardiac autonomic activity. In general, a decrease in respiratory frequency is associated with an increase in the heart period. In time-domain features, mean HR and RR interval shows more variability between normal and apnea patient which is shown in Fig. 3. The mean RR interval of normal case is high compared to apnea patient, because during apnea depolarization of ventricle is low compared to normal case.

The heart rate variability analysis approach uses Fourier transforms. The HRV spectrum contains two major components: First is high-frequency (0.18–0.4 Hz) HRV that has been used as an index of parasympathetic activity and is consecutive with respiration which is same as the respiratory sinus arrhythmia (RSA). The second is a low-frequency (0.04–0.15 Hz) component that appears to be mediated by both the vagus and cardiac sympathetic nerves. The area below the relevant frequencies presented in absolute units (square milliseconds) is the power of spectral components. The total power of a signal is equal to the variance of the entire signal, integrated over all frequencies. Parasympathetic–sympathetic balance is used as the ratio of the low-to-high frequency spectra [11]. In the control group, high-frequency (HF) band power is higher than apnea group which is shown in Fig. 4.

The time-domain feature SD1 shows higher variation between normal and apnea patient which is shown in Fig. 5. SD1 of normal case is high as compared to apnea patient. In case of heart rate variability (HRV), SD1 reveals both short- and long-term variations of the signal which is a visual pattern of the RR interval data [16].

The overall classification accuracy using this algorithm is 90.5%. Data sample size is small. In future, including larger dataset and use of different algorithms may increase the efficiency. In order to increase the accuracy of different methods like principal component analysis [5], Rusboost along with TQWT can be used for the efficient classification of OSA [4].

5 Conclusions

In this work, sleep apnea is detected from more commonly available physiological signal such as electrocardiogram (ECG). Some effective features are extracted from ECG signal. In time domain, the features such as RR interval mean, mean HR, NN50, and SDNN are extracted from ECG signal. In frequency domain, the features such as VLF power, HF power, LF power, standard deviation 1, and standard deviation 2 are extracted from ECG signal. Among those features, HF power (%), mean HR (bpm), standard deviation 1 (ms) show higher variation between normal and apnea ECG signals. The neural network classifier is used to classify an ECG signal normal data and apnea data on the basis of above features. The overall classification efficiency is 90.5%. In the future work, inclusion of large dataset and the use of different algorithms may increase the classification efficiency higher.

References

1. Rotariu C, Cristea C, Arotaritei D, Bozomitu RG, Pasarica A (2016) Continuous respiratory monitoring device for detection of sleep apnea episodes. In: 2016 IEEE 22nd international symposium for design and technology in electronic packaging (SIITME), pp 106–109
2. Martín-González S, Navarro-Mesa JL, Juliá-Serdá G, Kraemer JF, Wessel N, Ravelo-García AG (2017) Heart rate variability feature selection in the presence of sleep apnea: an expert system for the characterization and detection of the disorder. Comput Biol Med 91:47–58
3. Fan SH, Chou CC, Chen WC, Fang WC (2015) Real-time obstructive sleep apnea detection from frequency analysis of EDR and HRV using Lomb Periodogram. In: 2015 37th annual international conference of the IEEE engineering in medicine and biology society (EMBC), pp 5989–5992
4. Hassan AR, Haque MA (2017) An expert system for automated identification of obstructive sleep apnea from single-lead ECG using random under sampling boosting. Neurocomputing 235:122–130
5. Saxena N, Shinghal K (2015) Extraction of various features of ECG signal. Int J Eng Sci Emerg Technol 7(4):707–714
6. Rodríguez R, Mexicano A, Bila J, Cervantes S, Ponce R (2015) Feature extraction of electrocardiogram signals by applying adaptive threshold and principal component analysis. J Appl Res Technol 13(2):261–269
7. Mannurmath JC, Raveendra M (2014) MATLAB based ECG signal classification. Int J Sci Eng Technol Res (IJSETR) 3(7):1946–1950
8. Corrales MM, de la Cruz Torres B, Esquivel AG, Salazar MAG, Orellana JN (2012) Normal values of heart rate variability at rest in a young, healthy and active Mexican population. Health 4(07):377
9. Ma HT, Liu J, Zhang P, Zhang X, Yang M (2015) Real-time automatic monitoring system for sleep apnea using single-lead electrocardiogram. In: TENCON 2015-2015 IEEE region 10 conference, pp 1–4
10. QRS Detection Using Pan-Tompkins algorithm. https://archive.cnx.org/contents/611d4152-cf7f-4344-b56d-15775f92fbac@1/qrs-detection-using-pan-tompkins-algorithm
11. Khalil MM, Rifaie OA (1998) Electrocardiographic changes in obstructive sleep apnoea syndrome. Respir Med 92(1):25–27
12. BIOPAC Systems, Inc., https://blog.biopac.com/respiration-and-ecg/
13. Quiceno-Manrique AF, Alonso-Hernandez JB, Travieso-Gonzalez CM, Ferrer-Ballester MA, Castellanos-Dominguez G (2009) Detection of obstructive sleep apnea in ECG recordings using time-frequency distributions and dynamic features. In: Annual international conference of the IEEE engineering in medicine and biology society, 2009, EMBC 2009, pp 5559–5562
14. de Chazal P, Penzel T, Heneghan C (2004) Automated detection of obstructive sleep apnoea at different time scales using the electrocardiogram. Physiol Meas 25(4):967
15. Kocak O, Bayrak T, Erdamar A, Ozparlak L, Telatar Z, Erogul O (2012) Automated detection and classification of sleep apnea types using electrocardiogram (ECG) and electroencephalogram (EEG) features. In: Advances in electrocardiograms-clinical applications. Turkey
16. Behbahani S, Moridani MK (2015) Non-linear Poincaré analysis of respiratory efforts in sleep apnea. Bratisl Lek Listy 116(7):426–432

Performance Comparison of SVM Classifier Based on Kernel Functions in Colposcopic Image Segmentation for Cervical Cancer

N. Thendral and D. Lakshmi

Abstract Cervical cancer is the second most common cancer affecting women worldwide. It can be cured in almost all patients if detected and treated in time. Pap smear test has been broadly used for detection of cervical cancer. The conventional Pap smear test has several shortcomings including subjective nature, low sensitivity, and frequent retesting. In order to overcome this issue, colposcopy method is used for visual inspection of cervix with the aid of acetic acid and with proper magnification, abnormal cells to be identified. Thus, we propose a method for automatic cervical cancer detection using segmentation and classification. In this work, several methods used for detecting cervical cancer is discussed which uses different classification techniques like K-means clustering, texture classification and Support Vector Machine (SVM) to detect cervical cancer. The proposed work compares and determines accuracy for five types of kernel functions, namely Polynomial kernel, Quadratic kernel, RBF kernel, linear kernel, and Multi-Layer Perceptron kernel. Analysis shows that Multi-layer Perceptron kernel in SVM classifier provides the best performance with an accuracy of 98%.

1 Introduction

Cervical cancer is a cancer arising from the cervix. It is due to the abnormal growth of cells that have the ability to invade or spread to other parts of the body. Cervix region is made up of three types of tissues, namely Columnar Epithelium (CE), Squamous Epithelium (SE), and Aceto White (AW) region. Cervical cancer typically develops from precancerous changes from 10 to 20 years. About 90% of cervical cancer cases

N. Thendral (✉)
Faculty of Information and Communication, Anna University, Chennai, India
e-mail: thendral2082@gmail.com

D. Lakshmi
Department of Electronics and Communication, St. Joseph's College of Engineering, Chennai, India
e-mail: lakhramdevan@gmail.com

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_168

are squamous cell carcinomas, 10% are adenocarcinoma, and a small number are other types. Early on, typically no symptoms are seen. Later, symptoms may include abnormal vaginal bleeding, pelvic pain, or pain during sexual intercourse. Diagnosis is typically by cervical screening followed by a biopsy [1]. Medical imaging is then done to determine whether or not the cancer has spread. Several methods are available for diagnosis of cervical cancer. Some of them are Pap test, LCB test, and colposcopy test.

The Papanicolaou test (abbreviated as Pap test, also known as Pap smear, cervical smear, or smear test) is a method of cervical screening which is used to detect potentially precancerous and cancerous processes in the cervix (opening of the uterus or womb) [2]. A Pap smear is performed by opening the vaginal canal with a speculum, and then collecting cells at the outer opening of the cervix at the transformation zone. The collected cells are examined under a microscope to look for abnormalities. Pap test results report a normal (negative) when all the cells are of healthy size and shape. It reports an abnormal (positive) test if there are any changes in size and texture. An abnormal Pap test does not mean that women patients have cancer as the observations might tend to go wrong due to improper visualization. Pap test method is simple in operation but it takes more time to give the observation for a patient as it requires investigation in clinical laboratory. The next manual screening method is the Liquid-Based Cytology (LCB) test used for detecting cervical cancer uses 5% acetic acid in the biopsy of the cervical tissues which changes the AW region into white color is the way of diagnosing cervical cancer.

On the other hand, colposcopy is a method which uses a device called as colposcopy for visual inspection of cervix with the aid of acetic acid. Colposcopy is a microscope-like device used by doctors to examine women external genital area (vulna), vagina, and cervix with proper magnification so that abnormal cells are easily identified. Images obtained with colposcopy referred to as colposcopy images or cervical images are widely used as diagnostic and screening tool for cervical cancer. The accuracy of colposcopy is highly dependent on the physician's individual skills. In expert hands, colposcopy has been reported to have a high sensitivity (96%) and low specificity (48%) when differentiating abnormal tissues. These manual screening methods suffers from accuracy and is also time-consuming. So, the computerized methods of classification are used for detecting normal and abnormal cells [3]. The most common classification methods are Support Vector Machine (SVM), k-means clustering and texture classification. The input image is fed where the preprocessing stage is performed by using filters to eliminate unwanted noise from the image and to enhance its quality [4]. This image is used for feature extraction so as to extract the feature of nucleus, cytoplasm and background. After the features are extracted, the nucleus and cytoplasm are segmented from the image. Finally, the image is classified as normal and abnormal cells.

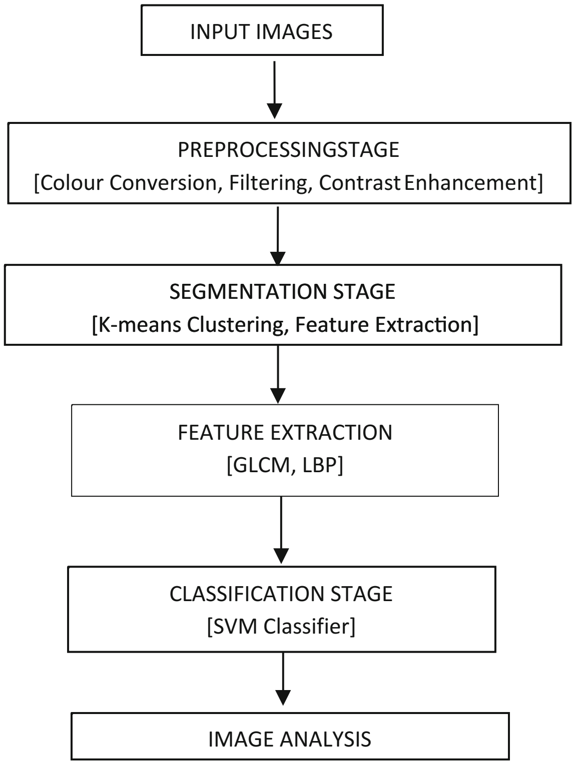


Fig. 1 Block diagram of the proposed system

2 Methodologies

The methodologies adopted in this paper are depicted in Fig. 1.

2.1 Preprocessing

Preprocessing of an image is the preparation of the sample to introduce it to an algorithm for the specified task. The aim of preprocessing is an improvement of image data that suppresses unwanted distortions or enhances some image features for further processing.

2.1.1 Color Conversion

Here, the RGB image is converted into gray scale image. Conversion of color image to gray scale image is one of the image processing applications used in different fields effectively. To convert the color images, the new algorithm performs RGB approximation, reduction, and addition chrominance and luminance. The gray scale conversion images generated using the algorithm preserved the salient features of the color image such as contrasts, sharpness, shadow, and image structure.

By emitting a restricted combination of three colors (red, green, and blue), we are able to generate almost any detectable color. This is the reason color images are often stored as three separate image matrices; one storing the amount of red (R) in each pixel, one the amount of green (G), and one the amount of blue (B).

Gray scale images do not differentiate gives dark pixels and much light is perceived as bright pixels. When converting an RGB image to gray scale, RGB values are considered for each pixel and make output as a single value reflecting the brightness of that pixel.

2.1.2 Image Fılteratıon

Image filteration is carried out using median filter. Median filtering is a nonlinear method used to remove noise from images. It is widely used as it is very effective at removing noise while preserving edges. The median filter works by moving through the image pixel by pixel, replacing each value with the median value of neighboring pixels. It suppresses the noise or other small fluctuations in the image; equivalent to the suppressions of high frequencies in the frequency domain. For smoothing the image, Median Filter is used which is a nonlinear operation often used in image processing to reduce "salt and pepper" noise. A median filter is more effective than convolution when the goal is to simultaneously reduce noise and preserve edges.

2.1.3 Contrast Enhancement

Image enhancement technique makes an image easier to analyze and interpret. The range of brightness values present on an image is referred to as contrast. Contrast enhancement is a process that makes the image features stand out more clearly by making optimal use of the colors available on the display or output device.

2.2 Segmentatıon

2.2.1 *K*-Means Clusterıng

K-means clustering is a type of hard clustering algorithm. It belongs to unsupervised cluster analysis algorithm and achieves partition clustering method. It is a key technique in pixel-based methods, where pixel-based methods based on *K*-means clustering are simple and the computational complexity is relatively low compared with other region-based or edge-based methods, the application is more practical. In the K-means algorithm, it initially defines the number of clusters k. Then, k-cluster center are chosen randomly. The distance between the each pixel to each cluster centers is calculated. The distance is the simplified Euclidean function. A single pixel is compared to all cluster centers using the distance formula. The pixel is moved to that particular cluster, which has the shortest distance amongst all. Then the center is recalculated. Again, each pixel is compared to all centers. The process continues until the center converges.

2.3 Feature Extraction

Feature extraction involves reducing the amount of resources required to describe a large set of data. Features are obtained using Local Binary Pattern (LBP), Gray Level Co-occurrence Matrix (GLCM), and LBP counts several binary patterns that occur in the image. GLCM is a kind of statistical approach that uses homogeneity, contrast, energy, and correlation information from pixels.

2.3.1 Gray Level Co-Occurrence Matrix (GLCM) Method

Gray Level Co-Occurrence Matrix has proved to be a popular statistical method of extracting textural feature from images. According to co-occurrence matrix, Haralick defines 14 textural features measured from the probability matrix to extract the characteristics of texture statistics of remote sensing images. The features like energy, correlation, homogenity, and contrast are obtained using GLCM method.

2.3.2 Local Binary Pattern (LBP) Method

LBP is the particular case of the texture spectrum model. It has further been determined that when LBP is combined with the Histogram of Oriented Gradients (HOG) descriptor, and it improves the detection performance considerably on some datasets.

LBP method divides the examined window into cells (e.g., 16×16 pixels for each cell). For each pixel in a cell, compare the pixel to each of its eight neighbors (on its left-top, left-middle, left-bottom, right-top, etc.). Follow the pixels along a circle, i.e., clockwise or counterclockwise. Where the center pixel's value is greater than the neighbor's value, write "0". Otherwise, write "1". This gives an 8-digit binary number (which is usually converted to decimal for convenience). Compute the histogram over the cell of the frequency of each "number" occurring (i.e., each combination of which pixels are smaller and which are greater than the center). This histogram can be seen as a 256-dimensional feature vector. In the computation of the LBP histogram, the histogram has a separate bin for every uniform pattern, and all nonuniform patterns are assigned to a single bin. Using uniform patterns, the length of the feature vector for a single cell reduces from 256 to 59.

2.4 Classıfıcatıon

2.4.1 Support Vector Machınes (SVMs)

Support Vector Machine (SVM) was first heard in 1992, introduced by Boser, Guyon, and Vapnik in COLT-92. Support vector machines (SVMs) are extreme learning machines. It has a set of related supervised learning methods used for classification and regression. They belong to a family of generalized linear classifiers. In another term, Support Vector Machine (SVM) is a classification and regression prediction tool that uses machine learning theory to maximize predictive accuracy. Support Vector Machines can be defined as systems which use hypothesis space of a linear functions in a high dimensional feature space, trained with a learning algorithm from optimization theory that implements a learning bias derived from statistical learning theory [5].

2.4.2 Kernal Functıons in SVM

Kernel function implicitly maps data from its original space to a higher dimensional feature space. Kernel function parameter selection is one of the important parts of support vector machine modeling. We propose emphasizes the classification task with support vector machine with different kernel function. It has several kernel functions including linear, polynomial, radial basis function, quadratic kernel, and multi-layer perception. The following are the kernel functions which are used to find out the accuracy [5–7].

RBF Kernel

The radial basis function network is one approach which has shown a great promise in this sort of problems because of its faster learning capacity. It has better capability of approximation to underlying functions, faster learning speed, better size of network, and high robustness to outliers. A radial basis function produces a piecewise linear solution which can be attractive. The equation for RBF Kernel is given by

$$K(x, y) = \exp\left(-\frac{\|x - y\|^2}{2\sigma^2}\right) \tag{1}$$

where > 0 is constant which denotes kernel width.

Linear Kernel

A linear kernel is often recommended for text classification with SVM because text data has lot of features and is often linearly separable. In text classification, both the number of instances and features are large. The equation for linear kernel is

$$k(x, y) = X_i.Y_j \tag{2}$$

Polynomial Kernel

Polynomial is one of the most powerful functions that have been used in many fields of mathematics such as fitting and regression. Low-order polynomial is desired for their smoothness, good local approximation and interpolation. A polynomial mapping is a popular method for nonlinear modeling. The equation of polynomial kernel is

$$K(x, y) = (x^T y + 1)^p \tag{3}$$

where $p > 0$ is a constant that defines the kernel order.

Quadratic Kernel

A new quadratic kernel the support vector machine is introduced to separate the data nonlinearly. A quadratic decision of separating nonlinearly the date is used. The equation for quadratic kernel is

$$K(x, y) = (x^T y + 1)^2 \tag{4}$$

Multi-layer Perceptron

MLP network is trained with cervical colposcopic images database with original features and then only extracted features. Classification performance of MLP in two cases is calculated and analyzed with help of network measures such as classification accuracy, recall, precision, mean squared error, and time [8, 9].

$$K(x) = \frac{(\tanh(x) + 1)}{2} - \frac{1}{1 + \exp(-2x)} \tag{5}$$

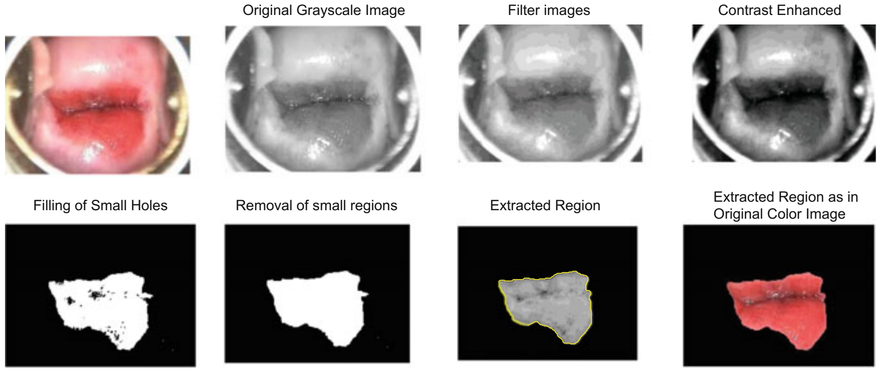
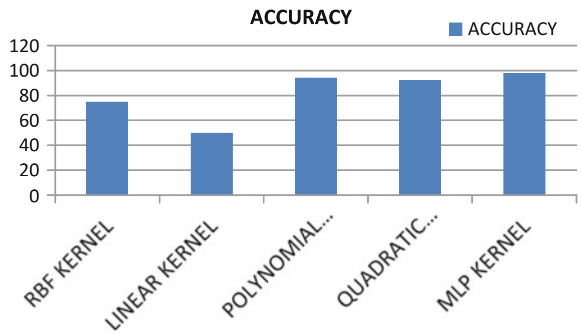


Fig. 2 Test image and its corresponding outcome

Fig. 3 Bar chart representation of accuracy by comparing five kernel functions



3 Results and Discussions

The extracted region of the affected area is identified using the proposed methodology for one sample image and is shown in Fig. 2.

The accuracy values of all the kernel functions is compared and represented in bar chart (Fig. 3).

The texture features, like contrast, correlation, energy, and homogenity are extracted from the segmented sample images by using GLCM are listed out in Table 1.

Finally, we compare the five types of kernels and the accuracy values of the kernel functions for the sample images are list out in Table 2.

4 Conclusion

Cervical cancer is one of the most death causing diseases around the world. Early detection is one of the most promising approaches in reducing the burden of growing cancer. In order to detect the cancer in early stage, we propose a method for auto-

Table 1 GLCM

Samples	Contrast	Correlation	Energy	Homogeneity
Sample 1	0.212779	0.97516	0.096202	0.916656
Sample 2	0.206783	0.973301	0.121533	0.924258
Sample 3	0.070206	0.982737	0.196052	0.966464
Sample 4	0.097927	0.984116	0.185056	0.954634
Sample 5	0.59759	0.992685	0.214383	0.974872
Sample 6	0.78818	0.989015	0.126878	0.964230
Sample 7	0.059680	0.991099	0.153083	0.975513
Sample 8	0.212045	0.956254	0.167082	0.951011
Sample 9	0.154843	0.974881	0.126620	0.933395
Sample 10	0.301584	0.947365	0.152555	0.942753

Table 2 Support Vector Machine techniques (values in %)

Samples	RBF kernel	Linear kernel	Polynomial kernel	Quadradic kernel	MLP
Sample 1	75	75	94	93	98
Sample 2	75	75	94	93	98
Sample 3	75	75	94	93	98
Sample 4	75	25	94	93	98
Sample 5	75	50	94	93	98
Sample 6	75	50	94	93	98
Sample 7	75	75	94	93	98
Sample 8	75	75	94	93	98
Sample 9	75	75	94	93	98
Sample 10	75	75	94	93	98

matic cervical cancer detection using segmentation and classification. The algorithms which are used in the four stages of preprocessing, segmentation, feature extraction, and classification are summarized [10, 11]. Another significant contribution of the proposed work is that it compares the five types of kernel function in SVM classifier. Analysis shows that MLP kernel in SVM classifier provides the best performance with an accuracy of 98%. Thus with the help of maximum accuracy, the cancer can be detected in early stage.

References

1. Kashyap D et al (2016) Cervical cancer detection and classification using Independent Level sets and multi SVMs. In: 2016 39th international conference on telecommunications and signal

processing (TSP), pp 523–528
2. Kaaviya S, Saranyadevi V, Nirmala M (2015) PAP smear image analysis for cervical cancer detection. In: 2015 IEEE International conference on engineering and technology (ICETECH), pp 1–4
3. Obukhova NA, Motyko AA, Kang U, Bae S-J, Lee D-S (2017) Automated image analysis in multispectral system for cervical cancer diagnostic. In: 2017 20th conference of open innovations association (FRUCT), pp 345–351
4. Njoroge E, Alty SR, Gani MR, Alkatib M (2006) Classification of cervical cancer cells using FTIR data. In: 2006 international conference of the ieee engineering in medicine and biology society, pp 5338–5341
5. Teeyapan K, Theera-Umpon N, Auephanwiriyakul S (2015) Application of support vector based methods for cervical cancer cell classification. In: 2015 IEEE international conference on control system, computing and engineering (ICCSCE), pp 514–519
6. Rajinikanth V, Dey N, Satapathy SC, Ashour AS (2018) An approach to examine magnetic resonance angiography based on Tsallis entropy and deformable snake model. Futur Gener Comput Syst 18:160–172
7. Rajinikanth V, Raja NSM, Kamalanand K (2017) Firefly algorithm assisted segmentation of tumor from brain MRI using Tsallis function and markov random field. Control Eng Appl Informatics 19(3):97–106
8. Karthick J, Lakshmi D (2016) A Sub-Threshold SRAM design for stability improvement in low power. Int J Emerg Trends Sci Technol 3(5):383–387
9. Muthumaheswaran N, Kabilamani P, Lakshmi D (2016) A high gain e-band power amplifier using 45 nm CMOS technology. Int J Emerg Trends Sci Technol 3(5):353–354
10. Lakshmi D, Roy S, Ranganathan H (2014) Automated texture based characterization of fibrosis and carcinoma using low-dose lung CT images. Int J Imaging Syst Technol 24(1):39–44
11. Lakshmi D, Roy S, Ranganathan H (2014) ANOVA of texture based feature set for lung tissue characterization using CT Images. J Comput Appl 7(1):1–5

GPU Based Denoising Filter for Knee MRI

Shraddha Oza and Kalyani R. Joshi

Abstract MRI is a popularly used technique for diagnosing muscle and skeletal disorders, especially of the knee. For accuracy in diagnosis, the rician noisy knee image needs to be filtered using efficient denoising algorithm. In recent years, the spatial neighborhood bilateral filter is being explored by researchers for its capacity to retain edges and tissue structures. It is noted that increase in image resolution slows down performance of the bilateral filter effectively discouraging its use. The research work proposes a cost-effective accelerated solution to the problem by implementing CUDA-based bilateral filter as applied to T2-weighted sagittal knee MRI slice. The work suggests use of GPU shared memory for optimized implementation and better speedup. The speedup achieved for 3.96 Mpixel knee MR image is 114.27 times more than that of its CPU counterpart. The results indicate average occupancy of 90.15% for image size of 630^2 pixels, indicating effective parallelization. Also, over varying rician noise levels, the average PSNR achieved is 21.83455 dB indicating good filter performance.

1 Introduction

Magnetic Resonance Imaging (MRI) is one of the leading imaging modalities that employs strong magnetic field along with radio frequency pulses to generate images of any internal organ of the body. The noninvasive diagnostic technique is frequently used to investigate cartilage or muscle tear in the knee leading to decisions of appropriate treatment. The MR image of knee provides high-resolution images from all possible angles of bones, ligaments, cartilages, and muscles in the knee joint. The

S. Oza (✉)
E&Tc Department, Army Institute of Technology, Pune, India
e-mail: sdoza@aitpune.edu.in

K. R. Joshi
E&Tc Department, PES Modern College of Engineering, Pune, India
e-mail: krjpune@gmail.com

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_169

MRI technique may also help to effectively articulate cartilage deterioration and thus avoid unnecessary knee arthroscopy which is generally painful to patient [1, 2].

The process of MR imaging introduces a signal-dependent rician noise which is multiplicative in nature. The noise hinders visibility of the tissue structures in the image adversely affecting accuracy of diagnosis. In recent years, bilateral denoising filter has become popular with the researchers, due to its edge preserving ability of tissues and simplicity of implementation. The filter was proposed by Tomasi and Manduchi in the year 2005 and recently has found significant place in the domain of medical image processing [3–7]. The only disadvantage of the filter is its computational complexity which increases with increase in the image size and resolution. The MR imaging technique being digitized generates scans of knee joint with high resolution of the order of megapixels [8, 9]. The performance of bilateral filter slows down for such images which discourages its use for medical images. The research work proposes a solution in the form of accelerated bilateral filter using NVIDIA's CUDA GPU. The GPU devices are basically single-instruction multiple data architectures, thereby making them suitable for implementing image processing applications [10–12].

The GPU devices are extremely powerful parallel architectures which make it possible to achieve large concurrency and speedup. In the present research work, the bilateral filter is implemented on CUDA GPU GTX 830 M which is a blend of Maxell as well as Kepler architecture. It has a compute capability of 5.0 and in the present work, it is hosted by Intel i5 Quad-Core Processor @ 2.2 GHz with 12 GB RAM. Using this system, speedup achieved is 114.27 for a knee T2-weighted image slice with a resolution of 630 $\times$ 630. In the proposed work, the filter performance for rician noisy knee MRI is quantified using PSNR metric. For a maximum noisy image slice, the PSNR value achieved is 16.9 dB while is observed to be 27.88 dB for a minimum noisy image with $s = 0.1$ indicating good filter performance.

The sections are distributed as follows. Section 2 briefly gives details of knee MRI. Section 3 elaborates bilateral filter theory. Section 4 discusses CUDA programming model. Section 5 gives details of experimentation and results specific to implementation on GTX 830 M GPU device, while Sect. 6 states the conclusions of the work done.

2 Knee MRI

Magnetic resonance imaging (MRI) of the knee is performed to identify and analyze problems in the knee joint, such as bone tumors, damaged cartilage or ligaments, and arthritis. There exist different types of MRI sequences out of which T1-weighted and T2-weighted are most commonly used. The T1-weighted image is characterized by short echo time (TE) and pulse repetition time (TR) and is used to evaluate anatomy of the knee. In this image sequence, fatty bone marrow appears bright while the knee joint fluid appears dark. The reverse is true for T2-weighted and is more sensitive to pathological information of the knee joint internals. The choice of the image sequence depends on the type of disease or problem of the patient's knee [2, 3].

The MR image is corrupted with noise that is introduced from the molecular movement within body parts and the electrical resistance of the RF receiver coils. The MR image generated needs to have good quality with high signal-to-noise ratio for accurate diagnosis. MRI machine with higher magnetic field strength (>3 T) is able to generate images with higher SNR as well as of higher resolution in minimum time. In addition to magnetic field strength, relaxation time, TE/TR, and field of view are some of the key parameters affecting SNR of the MR Image. The noise in the MR image is modeled to be rician which is signal-dependent noise and is multiplicative in nature. The probability distribution of this noise appears to be rayleigh for SNR < 2 and is close to gaussian noise for SNR > 2 [3, 13, 14]. The present work uses T2-weighted image of the knee joint to build Rician noisy database for experimentation. The details of the same are discussed in Sect. 5. The next section briefly discusses the bilateral filter.

3 Bilateral Filter

Bilateral filtering is appropriate denoising algorithm for a Rician noisy image as it has the ability to retain contours. It works on the assumption that a pixel may influence another pixel in the neighborhood as well as the one which may not be in the neighborhood but anywhere in the image which has similar intensity value. The filter uses a spatial Gaussian which accounts for spatial proximity of pixels and a range Gaussian to identify photometric proximity between the pixels. The filter is defined as

$$BF[Ix] = \frac{1}{Wp} \sum_{y \in R} G_{\sigma_s}(\|x - y\|)\, G_{\sigma_r}(|Ix - Iy|) Iy \tag{1}$$

Here, R is a spatial neighborhood of pixel x. Spatial Gaussian weight G_{σ_s} is minimum for distant pixels, while G_{σ_r} is the range gaussian which is minimum for two dissimilar pixels. Filter performance is controlled by the values of standard deviation thresholds of space σ_s and range σ_r. In executing the filter equation, computations are done on each pixel with reference to all the pixels in the search area within the image. Thus, the computational complexity increases with the image and size of the neighborhood R. This also increases execution time of the filter which is unwanted. Therefore, bilateral filter may be discouraged to use for MR images with high resolutions as its performance would slow down [15–17]. The problem can be solved by porting the filter implementation to parallel domain. CUDA GPU avails an extensively parallel SIMD architecture on desktops and laptops making itself a cost-effective solution [18, 19]. The next section briefly discusses programming model of the CUDA GPU and implementation of bilateral filter using the same.

4 CUDA Programming Model

CUDA GPU was launched by NVIDIA Corporation in 2007 for general-purpose computing, and since then it has revolutionized parallel programming. Though parallel architectures like Cray, IBM cell were available, they required specialized programmers to use them. Also, GPUs were used especially for graphics applications and required OpenGL, OpenCL programming expertise. With CUDA compliant GPU, a professional C programmer could easily develop an application using CUDA C library which merely is an extension of C which brought the breakthrough.

CUDA compliant GPU architecture consists of multiple streaming multiprocessors (SMP), wherein each SMP has hundreds of cores (Fig. 1). A unique programming entity thread is mapped to a core and 32 such concurrently running threads are identified as a warp. A group of 16 or 32 warps are identified as a block, and a matrix of such blocks is defined as grid. In CUDA C programming environment, the kernel function is mapped to a grid of blocks which executes the function by launching many concurrent threads being executed by the cores. The CUDA programmer needs to predefine the concurrency in terms of number of blocks and threads with reference to the image matrix size. This concurrency factor defines the speedup achieved [20–22].

The data bandwidth between CPU memory and GPU memory is limited by that of PCI bus specification and often is a bottleneck in overall performance of GPU. Thus, these data transfers should be kept to a minimum and the GPU on-chip memory model should be made use of to achieve maximum performance. The GPU registers allocated per thread make up for the fastest storage while access to device memory slows down the performance speedup. The device allocates fast shared memory per block which when used optimizes the memory access to a great extent. The architecture also includes cached texture and constant memory which is read only and is of great use to accelerate the speed up.

The research work presented here uses shared memory to achieve the required speedup for bilateral filter implementation. It is one of the effective memory optimization techniques used to achieve efficient GPU performance [22]. The details are briefly discussed in Sect. 5.

Performance of GPU is generally quantified using occupancy index, number of active warps, branch diversions, and memory access. The occupancy index indicates the on-chip resources utilized by the kernel in terms of number of blocks instantiated, registers, and shared memory used [23]. For any CUDA application, these details are made available by the profiling tools such as NSight Visual Profiler. The details of experimentation and the profiler tool are discussed in the next section.

5 Experimentation

In the present work, the bilateral filter (BF) is implemented on CUDA GTX 830 M device (Fig. 2) hosted by Intel i5 Quad-Core processor @ 2.2 GHz with 12 GB RAM.

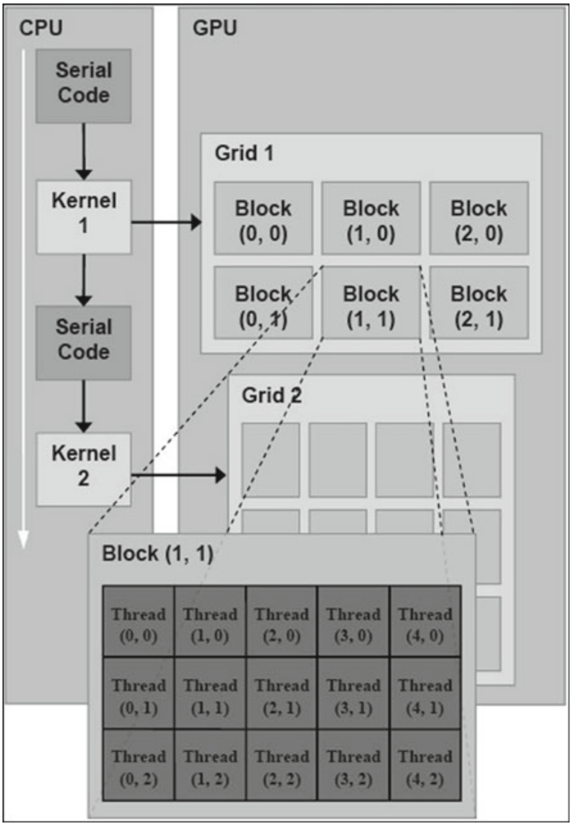


Fig. 1 CUDA GPU programming model

For building the test image database, T2-weighted sagittal knee MR image of 630 × 630 resolution is corrupted with varying levels of Rician noise from $s = 0.02$–0.8 in steps. The BF is applied to each test image, and the filter performance parameters such as MSE and PSNR are noted (Table 1). For quantifying the GPU performance, the execution time and memory access time for each test image are noted down along with number of blocks and occupancy index.

The proposed BF implementation uses shared memory which is very fast next to register access and is shared by all the threads in a block. In the present work, a tile of input MR image is loaded in shared memory for BF convolution. The scheduled threads are synchronized, and each thread operates on pixels in shared memory concurrently. On completion of computations, the tile is written back to global memory. This method accelerates the speedup by optimizing memory access. The observations of execution time and memory access are given in Table 1.

The CUDA C filter implementation uses OpenCV library for image read and display/write functions. OpenCV being an open-source image toolkit is extremely optimized and thus contributes to the overall speedup [24]. The filter radius is defined to be 5 × 5 and the filter parameters, namely, sigma domain and sigma range for BF

Fig. 2 Device properties of GTX 830M GPU

Cuda Device ID [Address] Device Process	[0] [0x0] GPU 0 - GeForce 830M T13.exe [12344]
MAX_THREADS_PER_BLOCK	1024
MAX_BLOCK_DIM_X	1024
MAX_BLOCK_DIM_Y	1024
MAX_BLOCK_DIM_Z	64
MAX_GRID_DIM_X	2147483647
MAX_GRID_DIM_Y	65535
MAX_GRID_DIM_Z	65535
MAX_SHARED_MEMORY_PER_BLOCK	49152
TOTAL_CONSTANT_MEMORY	65536
WARP_SIZE	32
MAX_PITCH	2147483647
MAX_REGISTERS_PER_BLOCK	65536
CLOCK_RATE	1150000
TEXTURE_ALIGNMENT	512
GPU_OVERLAP	1
MULTIPROCESSOR_COUNT	2
KERNEL_EXEC_TIMEOUT	1
INTEGRATED	0
CAN_MAP_HOST_MEMORY	1
COMPUTE_MODE	0
MAXIMUM_TEXTURE1D_WIDTH	65536
MAXIMUM_TEXTURE2D_WIDTH	65536
MAXIMUM_TEXTURE2D_HEIGHT	65536
MAXIMUM_TEXTURE3D_WIDTH	4096

kernel (GPU), are varied from 20 to 200 with increase in noise level so as to achieve optimum filter performance. The readings for PSNR and MSE are given in Table 1. The GPU implementation is done with block size of 1024 threads and grid size of maximum {32, 32, 1} blocks. The observations for actual GPU utilization summary are given in Fig. 5 Also, the work here implements bilateral filter in sequential domain (BF_CPU) in C language using the same host i5 machine for analysis of speedup. The comparative execution time is shown in Fig. 3 graphically.

Table 1 Observations for BF_GPU filter performance in terms of MSE, PSNR, execution time, and memory access time, for 630 × 630 sized knee MRI with varying rician noise level from $s = 0.02$ to 0.8

Noise Level (s)	MSE	PSNR (dB)	Execution time (msec)	Memory access time (msec) Host➔Device	Memory access time (msec) Device➔Host
0.02	115.5874	27.5017	1.412	0.28774	0.27136
0.05	123.4603	27.2155	1.414	0.28774	0.27546
0.1	148.8829	26.4024	1.406	0.29552	0.24269
0.2	246.4508	24.2135	1.384	0.28803	0.27136
0.3	398.2268	22.1295	1.371	0.30634	0.27443
0.4	588.8214	20.431	1.356	0.13414	0.27034
0.5	804.6602	19.0747	1.353	0.13517	0.27341
0.6	1042.385	17.9505	1.345	0.25088	0.23654
0.7	1273.116	17.0821	1.347	0.25091	0.23552
0.8	1508.768	16.3446	1.351	0.10854	0.26112

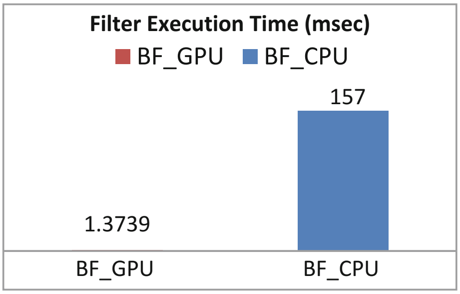


Fig. 3 Execution time of BF_GPU (1.3739 ms) versus BF_CPU (157 ms) showing 114.27 times speedup achieved

5.1 Observations and Discussion

It can be seen from Table 1 that, for minimum noise level of 0.02, the PSNR is 27.5 dB with MSE value as 115.5874. With increase in noise level from 0.02 to 0.8, the MSE increases to 1508.768 and PSNR drops to 16.3446 dB. The same is also shown graphically in Fig. 4a and b. The noisy images for $s = 0.02$, 0.5 and the respective filtered output images are given in Fig. 5 clearly indicating good filter performance.

The execution time with an average value of 1.3739 ms is achieved by GPU based bilateral filter while that achieved by its CPU counterpart is 157 ms indicating the GPU speedup of almost 114.27 (Fig. 3). The memory access time to copy the image from and to the CPU memory is tabulated in Table 1. The readings indicate total memory access time with an average value of 0.4957 ms which is almost one-third of the kernel execution time. This ratio is optimum because of shared memory being used.

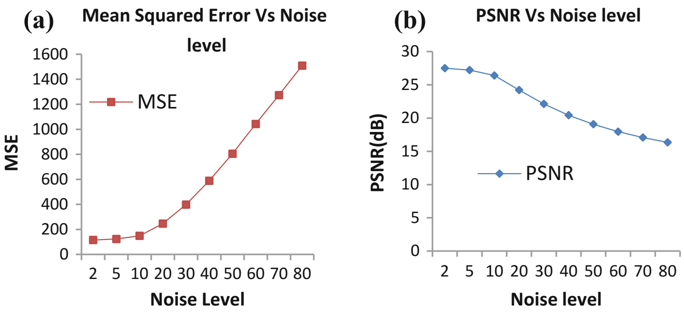


Fig. 4 Graphical representation of **a** MSE versus noise level, **b** PSNR (dB) versus noise level

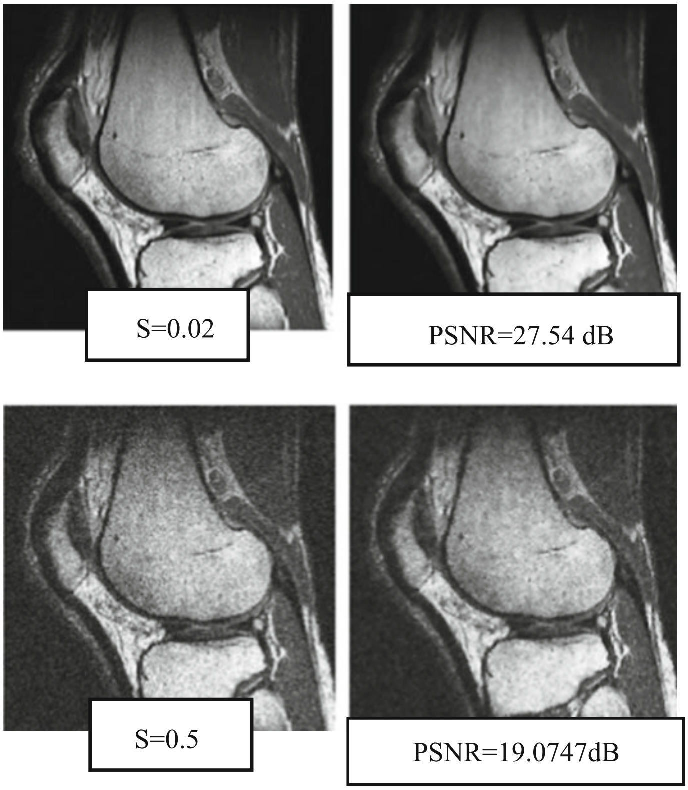


Fig. 5 Noisy image with s = 0.02, 0.5 and corresponding filtered output image with PSNR = 27.54 dB and 19.0747 dB, respectively

The observations from the report generated by nSight profiler are given in Fig. 6 and Table 2. It can be seen that the grid size, i.e., total number of blocks used, is 221(17, 13, 1) with 1024 threads being launched per block. The shared memory

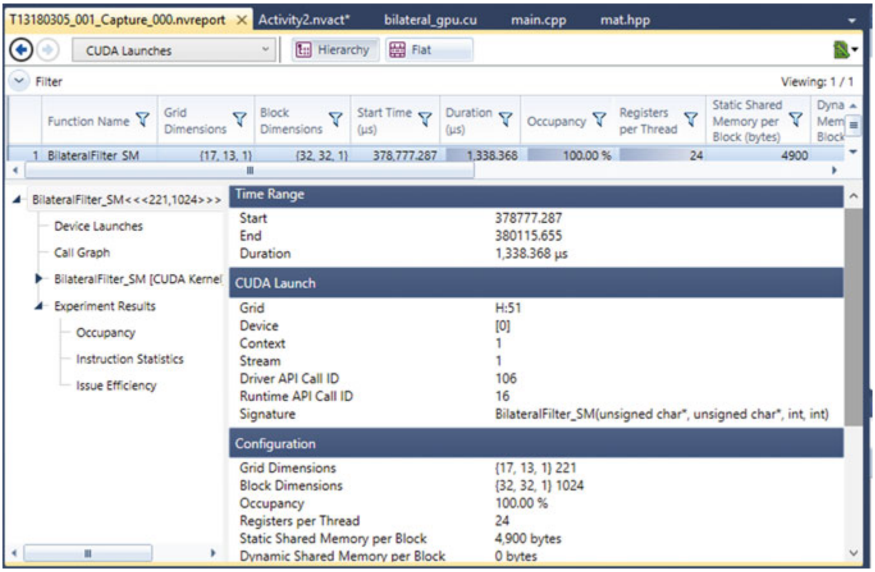


Fig. 6 Report of CUDA application BF profile generated by NSight visual profiler indicating execution time, memory access time, and device resource utilization in terms of blocks, threads, shared memory, and registers

Table 2 Observations of GPU device utilization for input image 630 × 630 pixels

Image Resolution (pixel s)	No. of Blocks/Grid	No. of Threads/Block	No. of Registers/Thread	Shared Memory used/Block (bytes)	Active warps (64 max)	Occupancy index (%)
630 × 630	221 {17,13,1}	1024	24	4900	64	90.15

used is 4900 bytes, and registers allocated per thread are 24. The usage of number of blocks and shared memory is unique to the test image size used here (630 × 630) and will change with change in image size. The number of active warps used for the filter implementation is 64 out of maximum 64 available (Table 2). In other words, the implementation is able to launch maximum possible number of threads indicating extremely efficient parallelization achieved. The occupancy index which is an index of utilization of resources, such as active warps, registers, and shared memory used, achieved here is 90.15%, again indicating the efficient parallelization.

6 Conclusion

The MRI of knee is significantly used by orthopedics for analyzing and diagnosing different problems such as knee pain, cartilage/muscle tear, and arthritis. The quality of the image needs to be with high SNR as it directly affects visibility of structures and tissues internal to knee. With digitization, the MR image size has increased to megapixels, thereby discouraging the use of bilateral filter, an efficient denoising algorithm.

The work presented here provides a cost-effective accelerated solution in the form of GTX 830 M CUDA GPU-based bilateral filter for T2-weighted sagittal knee MRI. The GPU implementation here uses shared memory technique giving speedup of 114.27 times over its CPU implementation. The GPU performance is characterized by occupancy index of 90.15%, clearly indicating efficient parallelization. The denoising performance of the BF measured in terms of PSNR is average 21.8345 dB indicating efficient noise removal for Rician noise level varying from 0.02 to 0.8. The results obtained here clearly underline the contribution of GPU to fast denoising of knee MR image. Further, it can be said that the CUDA GPU certainly would lead the research in coming years in the medical imaging domain.

Acknowledgements The research work acknowledges the extensive support provided by the management of Army Institute of Technology, Pune, India and of the research center, College of Engineering, Pune, India.

References

1. https://www.slideshare.net/benthungoe/mri-knee-of-orthopedic-importance-67500963
2. Rodrigues MB, Camanho GL (2010) Mri evaluation of knee cartilage. Rev Bras Orthop 45(4):340–346
3. Joshi KR, Oza SD (2018) MRI denoising for healthcare In: Kolekar MH, and Vinod Kumar (eds) Biomedical signal and image processing in patient care. IGI Global, pp 65–85
4. Oza SD, Joshi KR (2016) Performance analysis of denoising filters for MR images. In: Chakrabarti A, Sharma N, Balas VE (eds) Advances in computing applications. Springer, Singapore, pp 86–96
5. Tomasi C, Manduchi R (1998) Bilateral filtering for gray and colour images. In: Proceedings of the international conference on computer vision. IEEE, pp 839–46
6. Dougherty G (ed) (2011) Medical image processing: techniques and applications. Springer-Verlag, New York
7. Mohana J, Krishnavenib V, Guoc Y (2014) A survey on the magnetic resonance image denoising methods. Biomedical Signal Processing and Control, Elsevier, vol 9, pp 56–69
8. Rakhshan V (2014) Image resolution in the digital era: notion and clinical implications. J Dent Shiraz Univ Med Sci 15(4):153–155
9. Scholl I, Aach T, Deserno TM et al (2011) Challenges of medical image processing. Comput Sci Res Dev 26(5), Springer-Verlag
10. Shi L, Liu W, Zhang H, Xie Y, Wang D (2012) A survey of GPU-based medical image computing techniques. Quant Imaging Med Surg 2(3):188–206. AME Publishing Company
11. Moulika S, Boonna W (2011) The role of GPU computing in medical image analysis and visualization. In: Proceeding of Medical Imaging, vol 7967, SPIE
12. Kalaiselvi T, Sriramakrishnan P, Somasundaram K (2017) Survey of using GPU CUDA programming model in medical image analysis. Inf Med Unlocked 9:133–144. Elsevier
13. Gravel P, Beaudoin G, De Guise JA (2004) A method for modeling noise in medical images. IEEE transactions on medical imaging 23(10)

14. Gudbjartsson H, Partz S (2008) The rician distribution of noisy MRI data. In: HHS public access manuscript, PMC
15. Paris S, Kornprobst P, Tumblin J, Durand F (2008) Bilateral filtering: theory and applications. Comput Graph Vision 4(1):1–73
16. Zhang M (2006) Bilateral filtering in image processing, a thesis. Beijing University of Posts and Communications
17. Lekan M (2009) Impact of bilateral filter parameters on medical image noise reduction and edge preservation, a thesis. University of Toledo, Health Sciences Campus
18. Agarwal D, Wilf S, Dhungel A, Prasad SK (2012) Acceleration of bilateral filtering algorithm for many core and multicore architectures. In: Proceedings of IEEE 41st international conference on Parallel Processing (ICPP), pp 352–359
19. Larsson J (2015) A case study of parallel bilateral filtering on the GPU. Master Thesis, Malardalen University
20. Nvidia CUDA C (2012) Programming guide version 4.2
21. CUDA C (2012) Best practices guide DG-05603-001_v5.0. Design guide
22. http://www.training.praceri.eu/uploads/tx_pracetmo/GPU_Optimisation.pdf
23. CUDA Kernel occupancy : a detailed outlook. Application note, einfochips
24. www.opencv.org

Performance Evaluation of Audio Watermarking Algorithms Using DWT and LFSR

Ramesh Shelke and Milind Nemade

Abstract Advancement in Internet technology has compelled encryption, copyright protection and authentication of audios, videos, images, and documents. Watermarking techniques have been widely employed in copyright protection and authentications of digital media. Digital image watermarking techniques are extensively explored and found suitable for copyright protection of images, whereas audio watermarking techniques are less studied and need extensive research due to superior hearing ability of the human beings. In this paper, we present audio watermarking technique using discrete wavelet transform (DWT) and linear feedback shift registers (LFSR). Watermark signal is obtained using LFSR technique before embedding into audio signal. Dispersion of maximum power spectral density property of LFSR is explored that makes it suitable as scramblers. LFSR does not require secret key for scrambling and descrambling of watermark signal. Sequences were embedded into the DWT coefficients of the audio signal. Experimental simulations were performed to evaluate the performance of the audio watermarking technique. Audio watermarking finds applications in the area of ownership protection, tamper detection and authentication of music, military communication, voice-activated machines, and robots.

R. Shelke (✉)
Faculty of Engineering, Pacific Academy of Higher Education and Research University, Udaipur, India
e-mail: mail2shelke@gmail.com

M. Nemade
Department of Electronics Engineering, K. J. Somaiya Institute of Engineering and Information Technology, Mumbai, India
e-mail: mnemade@somaiya.edu

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_170

1 Introduction

Tremendous development in Internet, networking, and digital communication technology has fueled rapid sharing of digital media such as images, audios, videos, and documents. It has resulted in extensive demand for copyright protection and authentication of digital media. Due to sophisticated signal and image processing techniques, manipulation and duplication of digital data are much easier. Also, transmitting information over the communication networks is vulnerable for malicious attacks, and hence networks are insecure. Therefore, it is very much necessary to protect copyright information using watermarking techniques. The objective of the watermarking technique is to hide some information such as name, logo, signature, ID number, etc. into the actual media file without significant change in the original file [1, 2]. Whenever necessary, the hidden information can be extracted from the media to prove ownership or copyright. Audio watermarking is more difficult than image watermarking due to superior audio perception of human beings. Imperceptibility, robustness, and security are some of the most important characteristics of audio watermarking techniques. Security means that the embedded watermark cannot be extracted, manipulated, and changed by the unauthorized person, whereas robustness means least damage to the embedded watermark under various signal processing attacks and imperceptibility means embedded watermark must be inaudible. Furthermore, the audio watermarking technique must be computationally less complex and minimum time for embedding and extraction.

Many algorithms have been proposed for audio watermarking, and several of them are found to be too complex or require significant time for embedding/extraction or vulnerable to several forms of attacks [3]. However, complete security to the watermark signal is desirable without measurable distortion in the original audio signal. Patchwork techniques were proposed for image watermarking and extended further for audio watermarking. Patchwork techniques employ discrete cosine transform (DCT) and Fourier transform (FT) coefficients of original audio signals to obtain patches [4, 5]. Mostly, these patches do not have similar statistical properties. Similar technique was explored using wavelet transform coefficients to obtain two patches per audio segment. But it was unsuitable since many such audio patches do not have similar statistical properties. Hence, few patches were selected for watermarking. It also became difficult to identify that which patches were watermarked. The watermark embedding and extraction were explored using spread spectrum (SS) technique. SS sequence can be added to the host audio samples in time domain [5], FFT coefficients, in sub-band domain [6, 7], cepstral coefficients [8], and in a compressed domain [9]. Watermark signal is scrambled using LFSR technique before embedding into audio signal. Dispersion of maximum power spectral density property of LFSR is explored that makes it suitable as scramblers [10].

In this paper, performance of audio watermarking algorithms using DWT and linear feedback shift registers (LFSR) is evaluated to determine its suitability for audio watermarking techniques. Watermark signal is obtained using LFSR technique before embedding into audio signal. Dispersion of maximum power spectral density property of LFSR is explored that makes it suitable as scramblers. LFSR does not require secret key for scrambling and descrambling of watermark signal. Sequences were embedded into the DWT coefficients of the audio signal. The paper is organized as follows: Sect. 1 introduces to watermarking; Sect. 2 describes watermarking embedding and extraction algorithm in details. Results are discussed in Sect. 3 and finally concluded in Sect. 4.

2 Watermark Embedding and Extraction Algorithm

In this section, watermark embedding algorithm is discussed along with extraction algorithm. Extraction is exactly the reverse process of embedding.

2.1 Embedding Algorithm

Let *x[n]* and *w[n]* represent the original input audio signal and watermark signal, respectively, with m number of samples. The steps involved in the embedding of the audio signals are as follows:

1. Read *x[n]* of the length m from the input audio signal.
2. Characteristic polynomial is

$$p(x) = x^4 + x^3 + x + 1 \tag{1}$$

   a. Compute the sequences by shifting the elements in circular manner.
   b. The sequences are obtained as

$$p = \{p_1,\ p_2,\ p_3,\ ---,\ p_m\} \tag{2}$$

3. We apply DWT to input audio signal *x[n]*

$$\varphi(t) = \frac{1}{\sqrt{a}}\varphi\left(\frac{t-b}{a}\right) \tag{3}$$

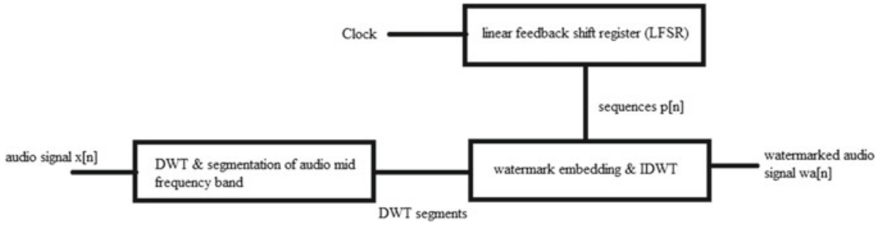


Fig. 1 Watermark embedding procedure

It is well known that very low-frequency and high-frequency components are vulnerable to signal processing attacks such as filtering and compression. Therefore, only mid-frequency range band can be used for watermark embedding.

4. We select the mid-frequency DWT coefficients for watermark embedding. These coefficients were divided into m segments of length N where m is the number of samples and N is the number of bits per watermark sequence sample.

$$a_i(k) = \{a_i(0),\ a_i(2), - - -, a_i(N-1)\} \tag{4}$$

where $i = 0, 1, 2, 3,—, m–1$

5. We have m number of sequences each with N bits for watermark embedding. We now embed watermark bit into the segments of DWT coefficients using (5).

$$wa_{i,k}(0) = \ a_{i,k}(0)\, x \text{ or } p_i(k) \tag{5}$$

where $i = 0, 1, 2, 3,—, m-1$ and $k = 0,1, 2,—N$-1

6. After obtaining all watermark segments, inverse discrete wavelet transform (IDWT) is applied. Figure 1 shows the watermark embedding procedure.

2.2 Extraction Algorithm

The algorithm extracts the watermark from the received audio signal, with the use of original sequence p_1. Let $y[n]$ is the received watermarked audio signal with or without signal processing attacks.

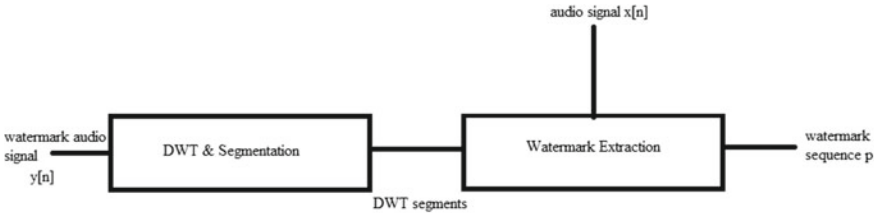


Fig. 2 Watermark extraction procedure

1. Read $y[n]$ of the length m from the input watermarked audio signal.
2. Characteristics polynomial is

$$p(x) = x^4 + x^3 + x + 1 \tag{6}$$

   a. Compute the sequences by shifting the elements in circular manner.
   b. The sequences starting from p_1 are obtained as

$$p = \{p_1,\ p_2,\ p_3, ---,\ p_m\} \tag{7}$$

3. We apply DWT to input audio signal *y[n]*

$$\varphi(t) = \frac{1}{\sqrt{a}}\varphi\left(\frac{t-b}{a}\right) \tag{8}$$

   Only mid-frequency range band was embedded with watermark; therefore, those coefficients are obtained.
4. We select the mid-frequency DWT coefficients for watermark embedding. These coefficients were divided into m segments of length N where m is the number of samples and N is the number of bits per watermark sequence sample.

$$ya_i(k) = \{ya_i(0),\ ya_i(2), ---, ya_i(N-1)\} \tag{9}$$

5. We have m number of sequence with N bits each. The watermark is now extracted from the segments of the DWT coefficients using (10)

$$p_i(k) = \ ya_{i,k}(0)\ xor\ a_{i,k}(0) \tag{10}$$

Figure 2 shows the watermark extraction procedure.

3 Experimental Results

Simulation results are depicted in this section using computer programming language MATLAB. Nine different audio signals were recorded using mobile phone with the sampling frequency of 44 kHz. Sound section was manually extracted from the audio, and inaudible section was removed. The audio signal was downsampled to obtain 4000 samples and quantized with 16 bits. Following types of attacks were used in the evaluation of the performance of the watermarking algorithm.

1. Low-pass filtering (LPF): low-pass filter with 12 kHz cut off frequency is applied to watermarked signal.
2. High-pass filtering (HPF): high-pass filter with 100 Hz cutoff frequency is applied to watermarked signal
3. MP3 attack (Compression): layer III compression was performed on watermarked signal.
4. Noise attack (NA): random noise of signal-to-noise ratio of 20 dB was added to watermarked signal.
5. Amplitude attack (AA): amplitude of the watermarked signal was increased by 1.25 times.

The robustness of the watermarking algorithm was measured using extraction rate (ET) given as (11)

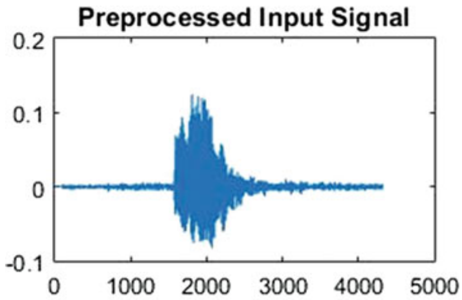
$$\mathrm{ET} = \frac{\text{correctly extracted watermark}}{\text{total embedded watermark}} \times 100\,(\%) \tag{11}$$

Table 1 depicts the measured ET for all the nine audio signals under various signal processing attacks. Figure 3 shows the three audio signals before watermark embedding.

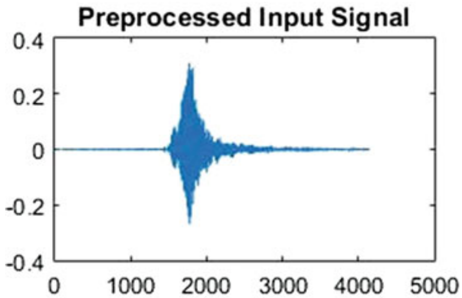
Table 1 Extraction rate

Audio signal/attacks	Extraction rate (ET) %								
	A1	A2	A3	A4	A5	A6	A7	A8	A9
LPF	91	92	94	93	94	92	90	91	93
HPF	96	94	95	97	96	94	93	95	94
MP3	82	81	83	81	83	83	80	81	83
NA	89	88	92	91	93	91	89	90	92
AA	98	97	99	98	99	98	96	97	98

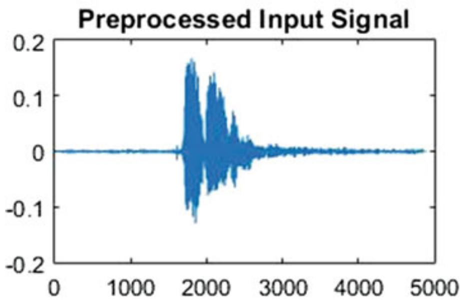
Fig. 3 Audio signals **a** A1, **b** A7, **c** A9



(a) A1



(b) A7



(c) A9

4 Conclusion

In this paper, performance of audio watermarking algorithms using DWT and linear feedback shift registers (LFSR) is evaluated to determine its suitability for audio watermarking techniques. Watermark signal is obtained using LFSR technique before embedding into audio signal. Experimental results clearly indicates suitable extraction rate for various types of attacks. Furthermore, extraction rate needs to be improved for MP3 or compression attack, LPF, and HPF. LFSR technique provides security to watermark signal by introducing randomness that makes it unpredictable.

References

1. Shanshan Li, Tingting Liu, Kun Bao, Ruijuan Qiao (2014) The study on the validity and reliability of digital audio watermarking. Comput Secur 2:34–36
2. Kob S, Nishimura R, Suzuki Y (2005) Time-spread echo method for digital audio watermarking. IEEE Trans Multimedia 7(2):212–221
3. Radha N, Venkatesulu M (2012) A chaotic block cipher for real-time multimedia. J Comput Sci 8(6):994–1000
4. Rahman MdM, Saha TK, Bhuiyan MdAA (2012) Implementation of RSA algorithm for speech data encryption and decryption. Int J Comput Sci Netw Sec 12(3):74–82
5. Kumar Singh, B, Tsegaye A, Singh J (2017) Probabilistic data encryption using elliptic curve cryptography and arnold transformation. In: IEEE international conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud), pp 644–651
6. Elshazly AR et al (2016) Synchronized double watermark audio watermarking scheme based on a transform domain for stereo signals. In: IEEE fourth international ja-pan-egypt conference on electronics, communications and computers
7. Balamurugan R, Kamalakannan V, Rahul Ganth D, Tamilselvan S (2014) Enhancing security in text messages using matrix based mapping and elgamal method in El-liptic curve cryptography. In: IEEE international conference on contemporary computing and informatics, pp 103–106
8. Xiang Y, Natgunanathan I, Peng D, Zhou W, Yu S (2012) A dual-channel time-spread echo method for audio watermarking. IEEE Trans Inf Forensics Sec 7(2):383–392
9. Kang H, Yamaguchi K, Kurkoski B, Yamaguchi K, Kobayashi K (2008) Full-index-embedding patchwork algorithm for audio watermarking. IEICE Trans Inf Syst, E91–D(11):2731–2734
10. Menezes AJ, van Oorschot PC, Vanstone SA (2001) Handbook of applied cryptography, 5th edn, CRC Press

Biometric Image Encryption Algorithm Based on Modified Rubik's Cube Principle

Mahendra V. Patil, Avinash D. Gawande and Dilendra

Abstract Multi-biometric systems are being mostly installed in many large-scale biometric applications most importantly UIDAI system in India. However, multi-biometric systems require storage of multiple biometric templates for each user that results in increased risk to user privacy and cybercrimes. Among all the biometrics, fingerprints are widely used in personal verification systems especially from small-scale to large-scale organization. Encryption of the biometric images is very important due to increasingly use of insecure systems and networks for storage, transmission, and verification. We proposed encryption algorithm based on modified Rubik's cube principle for fingerprint biometric images in this paper. Secret key for the encryption is obtained from the same fingerprint biometric image that assists in key management and storage for further decryption and verification process. The principle of Rubik's cube involves row and column rotation or changing of the pixels position. Further, bitwise XOR operation through modifications in secret key adds to security of the biometric images. Experimental simulations were performed to obtain the encryption parameters and validate the algorithm through common attacks such as noise and compression. Thus, the proposed encryption algorithm may be applied to fingerprint images used in attendance monitoring systems since it provides key management and does not involve complex calculations.

M. V. Patil (✉) · Dilendra
Faculty of Engineering, Pacific Academy of Higher Education and Research University, Udaipur, India
e-mail: apmahendra@yahoo.com

A. D. Gawande
Department of Computer Science and Engineering, Sipna College of Engineering and Technology, Amravati, India

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_171

1 Introduction

With the tremendous development in Internet and communication technology, sharing of audio, images, video, and biometric images has also increased. Users of the multimedia data are able to carry out real-time streaming video clips, video conferencing, listening to music, etc. However, most of the communication networks and systems are insecure and vulnerable to virus attacks. Therefore, these are unsuitable for communicating sensitive and important multimedia content such as defensed, commercial, or private information, and biometric images and videos. Thus, it necessitates the usage of encryption algorithm in the protection of valuable information.

Encryption algorithms have been extensively studied by the researchers for the data previously. This standard has resulted in algorithms like Data Encryption Standard (DES), Advanced Encryption Standard (AES), RC4, and RC6. However, application of these algorithms for images may not result in practically beneficial image encryption schemes. Image encryption schemes are different from data encryption schemes because image data are usually large in size, require more time for encryption, and also require a comparable amount of time for decryption. Computational overhead makes the traditional encryption algorithm inappropriate for real-time image encryption and decryption [1–3].

Multi-biometric systems are being mostly installed in many large-scale biometric applications most importantly UIDAI system in India. However, multi-biometric systems require storage of multiple biometric templates for each user that results in increased risk to user privacy and cybercrimes. Among all the biometrics, fingerprints are widely used in personal verification systems especially from small-scale to large-scale organization. Encryption of the biometric images is very important due to increasingly use of insecure systems and networks for storage, transmission, and verification. This provides a strong motivation for developing techniques that can protect the biometrics fingerprint data particularly during communication and transmission over insecure network channels. Many solutions based on spatial-domain and frequency- or transform-domain approach are proposed [4–7]. Spatial-domain technique uses traditional cryptography algorithms, whereas frequency-domain technique uses transforms such as Wavelet and Fourier. In both the cases, the basic idea is to shuffle pixel positions in the biometric images and chaotically generate transform orders [8–10].

Two important attributes of a powerful encryption technique are diffusion and confusion. Hence, the motivation is to generate a scheme which has more diffusion and confusion. Due to some intrinsic features of biometric images, some of the traditional encryption schemes are not very suitable for image encryption. Hence, the motivation is to generate schemes to encrypt images efficiently. To enhance the security of images, a biometric cryptosystem approach combines cryptography and biometrics. Under this approach, the biometric image is encrypted with the help of key. A key generated with the combination of fingerprint and password is used for image encryption. This mechanism is seen to enhance the security of biometrics images during storage, transmission, and verification.

We proposed encryption algorithm based on modified Rubik's cube principle for fingerprint biometric images in this paper. Secret key for the encryption is obtained from the same fingerprint biometric image that assists in key management and storage for further decryption and verification process. The principle of Rubik's cube involves row and column rotation or changing of the pixels position. Further, bitwise XOR operation through modifications in secret key adds to security of the biometric images. The paper is organized as follows: Sect. 1 introduces encryption; Sect. 2 describes encryption and decryption algorithm in details. Results are discussed in Sect. 3 and finally concluded in Sect. 4

2 Modified Rubik's Cube Algorithm

In this section, modified Rubik's cube encryption algorithm is discussed along with decryption algorithm. Decryption is exactly the reverse process of encryption.

2.1 Encryption Algorithm

Let I represent the 8-bit grayscale input biometric image of the size m x n. The steps involved in the encryption of the biometric image are as follows:

1. Generate two vectors k_r and k_c of the length m and n, respectively, from the input biometric images randomly from any one row and column.
2. Decide the number of iterations $i$max and initialized counter $i = 0$.
3. For each row x in I,
   a. Calculate the sum of all elements in the row x represented by ar(x)
   b. Compute modulo 2 of ar(x) represented by mar(x)
   c. Left or right circular shift each pixel elements $k_r(x)$ times according to the following if mar(x) = 0 right shift else left shift.
4. For each column y in I,
   a. Calculate the sum of all elements in the column y represented by ac(y)
   b. Compute modulo 2 of ac(y) represented by mac(y)
   c. Up or down circular shift each pixel elements $k_c(y)$ times according to the following if mac(y) = 0 upshift else downshift.
   Steps 3 and 4 will create a scrambled image represented by I_s.
5. Use vector k_r to apply bitwise XOR operator on each row of the partial resultant image I_s using the following equations:

$$I_1(x, y) = I_s(m - x, n - y) \text{XOR } k_r(x) \quad (1)$$

where XOR is the XOR operator applied on flipped scrambled image upside down with vector $k_r(x)$.

6. Use vector k_c to apply bitwise XOR operator on each column of the partial resultant image I_1 using the following equations:

$$I_2(x, y) = I_1(m - x, n - y)\text{XOR}\, k_c(y) \tag{2}$$

where XOR is the XOR operator applied on flipped partially encrypted image upside down with vector $k_c(y)$.

7. Increment the counter $i = i + 1$.
8. Repeat steps 3–7 if $i < i$max else stop.

The encrypted image Ie is obtained, and thus the process of encryption is completed. Vectors k_r and k_c are designated as secret key and obtained randomly from the input biometric image. Also, vectors k_r and k_c are randomly hidden into the final encrypted image and thus reduce key management. Vectors k_r and k_c can be again obtained from the encrypted image during decryption for further processing of biometric image.

2.2 Decryption Algorithm

Let Ie represent the 8-bit grayscale input encrypted biometric image of the size $m + 1 \times n + 1$ due to hidden vectors k_r and k_c. The steps involved in the decryption of the encrypted biometric image are as follows:

1. Obtain two vectors kr and kc of the length m and n, respectively, from the encrypted biometric image and restore the size of the encrypted image to $m \times n$.
2. Determine the number of iterations $i$max and initialized counter $i = 0$.
3. Use vector k_c to apply bitwise XOR operator on each column of the encrypted image I_e using the following equations:

$$I_{11}(x, y) = I_e(m - x, n - y)\text{XOR}\, k_c(y) \tag{3}$$

where XOR is the XOR operator applied on flipped encrypted image upside down with vector $k_c(y)$.

4. Use vector k_r to apply bitwise XOR operator on each row of the partial resultant image I_{11} using the following equations:

$$I_{21}(x, y) = I_{11}(m - x, n - y)\text{XOR}\, k_r(x) \tag{4}$$

where XOR is the XOR operator applied on flipped scrambled image upside down with vector $k_r(x)$.

5. For each column y in I_{21},

   a. Calculate the sum of all elements in the column y represented by ac(y)
   b. Compute modulo 2 of ac(y) represented by mac(y)
   c. Up or down circular shift each pixel elements $k_c(y)$ times according to the following if mac(y) = 1 upshift else downshift.

6. For each row x in I_{21},

   a. Calculate the sum of all elements in the row x represented by ar(x)
   b. Compute modulo 2 of ar(x) represented by mar(x)
   c. Left or right circular shift each pixel elements $k_r(x)$ times according to the following if mar(x)= 1 right shift else left shift.
   Steps 5 and 6 will create a final decrypted image represented by Id.

7. Increment the counter $i = i + 1$.
8. Repeat steps 3–7 if $i < i$max else stop.

The complete process of encryption and decryption with secret key management policy facilitates security of the biometric images during the process of transmission, storage, and verification. Also, the modified algorithm overcomes the disadvantage of not having constant values in vectors k_r and k_c. The complete process is based on changing pixel position in more confused manner with no secret key generation and management policy.

3 Experimental Results

Visual testing, entropy analysis, noise attack, and compression attack was performed on the obtained experimental results through modified Rubik's cube principle and algorithms.

3.1 Visual Testing

In this experiment, biometric fingerprint image of size 125 × 125 pixels was used for visual testing. Figure 1 illustrates the input biometric fingerprint image and its encrypted output image through the proposed algorithm. It is clearly understood that there is absolutely no pictographic similarity between these images. It is desired that the encrypted image should mostly be different from its original input image. Mostly two parameters are obtained to compute the differences between input and encrypted image. First, the measure is the number of pixels change rate (NPCR) that specifies the percentage of dissimilar pixels between two images. The second one is the unified average changing intensity (UACI), which measures the average intensity of dissimilarities in pixels between two images [11]. To achieve the performances of an ideal image encryption algorithm, NPCR values must be as large as possible, and

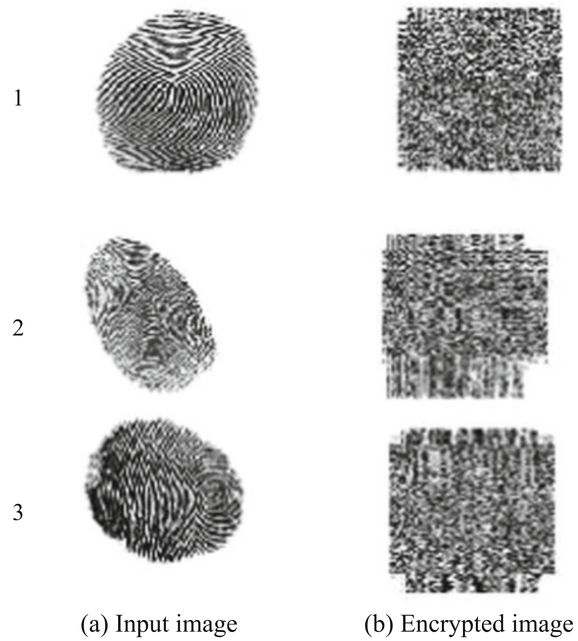


Fig. 1 Result of visual testing

Table 1 NPCR and UACI

Biometric images	NPCR	UACI
Fingerprint 1	99.18	25.02
Fingerprint 2	98.17	28.87
Fingerprint 3	98.98	25.32

UACI values must be around 30%. Table 1 indicates the obtained values of NPCR and UACI for the original input image and their encrypted counterpart. Thus, the values are very close to 100% for the NPCR measure and also the UACI values are also acceptable. The high percentage values of the NPCR measure clearly indicates that the pixels locations have been arbitrarily changed. Furthermore, the UACI values show that practically all pixel grayscale values of encrypted image have been altered from their values in the original input image.

3.2 Entropy Analysis

For 8-bit grayscale images having 256 levels, if each level of gray is assumed to be equiprobable, then the entropy of this image will be theoretically equal to 8 (no. of bits). Thus, it is understood that ideally an encryption algorithm for biometric images should give an encrypted image having entropy nearly equal to 8. Table 2 shows the entropy values of the three input fingerprint images and their respective encrypted images.

Table 2 Entropy values

Number	Input image	Encrypted image
Fingerprint 1	6.5	7.96
Fingerprint 2	5.5	7.89
Fingerprint 3	6.8	7.97

3.3 Noise Attack

To measure the robustness of the proposed modified Rubik's cube algorithm, random noise was added into the encrypted image. Thereafter, the noisy encrypted image was decrypted using algorithm. In this experiment, salt-and-pepper noise of density 0.01 was added into the encrypted image. Figure 2 shows the noisy input image and its encrypted and decrypted counterpart images. Results clearly indicate that noise attack such as salt-and-pepper noise has significant effect on the decrypted image.

Fig. 2 Result of noise attack

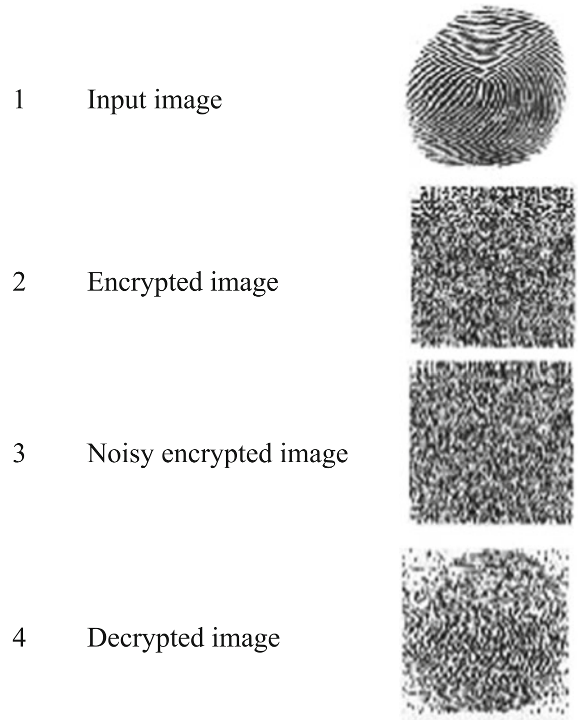
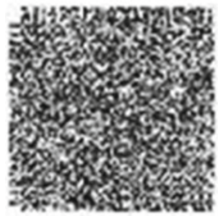


Fig. 3 Result of compression attack

1	Input image	
2	Encrypted image	
3	Decrypted image	

3.4 Compression Attack

Mostly, in the biometric image systems, the original input fingerprint image will be scanned, encrypted, and stored. Therefore, it is mandatory to evaluate the effect of compression on the decryption process. In this experiment, the original encrypted image was compressed using JPEG format and then decrypted. Figure 3 shows the input image, encrypted image, and decrypted image obtained after encoding and decoding using JPEG algorithm. Thus, it is clearly illustrated that compression does not significantly affect the decryption process and is robust against it.

4 Conclusion

We proposed modified encryption algorithm based on Rubik's cube principle for fingerprint biometric images in this paper. Secret key for the encryption is obtained from the same fingerprint biometric image that assists in key management and storage for further decryption and verification process. Experimental simulations were performed to obtain the encryption parameters and validate the algorithm through common attacks such as noise and compression. The simulation was done using MATLAB software on Intel i5 processor with 3 GHz speed and 4 GB RAM. Visual testing, entropy analysis, noise attack, and compression attack were performed on the obtained experimental results. The proposed algorithm is best suitable for biometric image encryption especially fingerprint. It is robust to compression attack and satisfies all encryption parameters such as NPCR, UACI, and entropy. Further, the robustness of the proposed algorithm needs to be improved for noise attack.

Acknowledgements Fingerprint images are used the author and co-authors. The Consent of all participants was taken.

References

1. Murugan B, Gounder AGN (2016) Image encryption scheme based on blockbased confusion and multiple levels of diffusion. IET Comput Vision 10(6):593–602
2. Bringer J, Chabanne H, Patey A (2013) Privacy-preserving biometric identification using secure multiparty computation: An overview and recent trends. Sig Process Mag 30(2):42–52
3. Bhatnagar G, Wu QJ, Raman B (2012) A new fractional random wavelet transform for fingerprint security. IEEE Trans Syst Man Cybern Part A Syst Hum 42(1):262–275
4. Bhatnagar G, Wu QJ (2012) Chaos-based security solution for fingerprint data during communication and transmission. IEEE Trans Instrum Meas 61(4):876–887
5. Norouzi B, Seyedzadeh SM, Mirzakuchaki S, Mosavi MR (2013) A novel image encryption based on row-column, masking and main diffusion processes with hyper chaos," Springer Multimedia Tools and Applications, pp 1–31
6. Nagar A, Nandakumar K, Jain AK (2012) Multibiometric cryptosystems based on feature-level fusion. IEEE Trans Inf Forensics Secur 7(1):255–268
7. Hongjuna L, Xingyuana W (2010) Color image encryption based on one-time keys and robust chaotic maps. Elsevier J Comput Math Appl 5:3320–3327
8. Philip Chen CL, Zhang T, Zhou Y (2012) Image encryption algorithm based on a new combined chaotic system. In: IEEE International conference on systems, man, and cybernetics
9. Wang Y et al (2011) A new chaos-based fast image encryption algorithm. Elsevier J Appl Soft Comput 11:514–522
10. Wu X, Wang Z (2015) A new DWT-based lossless chaotic encryption scheme for color images. In: IEEE international conference on computer and computational sciences
11. Chen G, Mao Y, Chui CK (2004) A symmetric image encryption scheme based on 3D chaotic cat maps. Chaos, Solitons Fractals 21(3):749–761

Leaf Disease Detection Based on Machine Learning

Anish Polke, Kavita Joshi and Pramod Gouda

Abstract Farming is a standout among the most imperative elements in light of which a nation's economy is chosen. Alignments in crops are very regular, which is one of the prominent factors that leads to the disease location and detection in plant's parts which is of high importance in agroindustry. In this way, it is imperative to effectively distinguish the maladies from the harvest to specifically shower herbicides and treat to diminish wastage utilization of concoction. In this work, we display an approach that coordinates picture handling and machine figuring out how to permit diagnosing infections from leaf pictures and to contemplate quantitative plant physiology imaging and PC vision which is utilized. Wavelet is extremely mainstream apparatus in picture preparing calculation. Surface highlights are utilized for recognition of yield. These features were mean, standard deviation, skewness, and kurtosis which we have used in this paper. We are proposing an approach, which used to identify the plant infection, i.e., plant disease. Here, we are using minimum distance classifier for the classification of the disease. The proposed approach displays a way toward robotized plant sicknesses finding on a gigantic scale.

1 Introduction

Indian economy is profoundly reliant of agrarian efficiency. One of the major factors responsible for the crop destruction is plant disease. Display strategy for infection discovery of plant is perception by specialists through their stripped eye which required

A. Polke (✉) · K. Joshi · P. Gouda
Department of Electronics and Telecommunication, G.H. Raisoni College of Engineering and Management, Pune, India
e-mail: polke_anish.ghrcemmevlsi@raisoni.net; anishpolke@gmail.com

K. Joshi
e-mail: kavita.joshi@raisoni.net

P. Gouda
e-mail: pramod.gouda@raisoni.net

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_172

a substantial group of specialists and also consistent observing of plant is required, which is expensive undertaking. In created nations, they are utilizing the bleeding edge advancements like data and space innovation, and biotechnology. It is important to change and use the exploration for eco-accommodating yield administration and expanded part of research-based innovations and supportable harvest generation. In plants, some broad sicknesses seen are dark colored and yellow spots, early and late singe, and others are parasitic, viral, and bacterial illnesses. Picture preparing [1] can be utilized for estimating influenced territory of ailment and to decide the distinction in the shade of the influenced zone. Programmed location of the illnesses by simply observing the indications on the plant leaves makes it simpler and additionally less expensive. Accordingly, in the field of farming, utilization of programmed illness recognition strategy in plants assumes a critical part. Recognition of a plant malady in extremely introductory stage would be exceptionally gainful. This additionally underpins machine vision to give picture-based programmed process control, examination, and robot direction. The straightforward standard of the activity of proposed framework is to take the reactions from different parameters that choose the profitability, handling them as per the calculation, and anticipate the appropriate yield for the land alongside the present status of product if tainted or not. Additionally, recommendation is given of a few manures that could be utilized to enhance the richness. For making framework accessible to average citizens, we propose that this framework is made as an Android application, where the agriculturist could bolster the information sources and get the essential outcome.

2 Literature Survey

After auditing the various reports, we can infer that the machine vision examined differs broadly in relying upon the intricacy of the visual scene. With the progression of the PC, the greater part of the on-field machine vision recognition work has been performed and come about, that is, beginning from singular leaves kept on plain foundation to limit leaves scenes of field plants developing in on-field characteristic conditions. Surfaces and size component are utilized for weed discovery and mechanization of manure splashing [2]. Wavelet change is utilized to extricate the surface highlights of the yield and weed pictures for order [3].

Singh and Misra [4] present the review on various diseases sorting methods utilized in plant leaf illness location and a calculation for picture division strategy that can be utilized for programmed identification and additionally characterization of plant leaf ailments. With less computational endeavors, the ideal outcomes were acquired, which likewise demonstrates the proficiency of proposed calculation in acknowledgment and characterization of the leaf illnesses. Another favorable position of utilizing this technique is that the plant's illnesses can be distinguished at the beginning period or the underlying organize. To enhance acknowledgment rate in grouping process, simulated neural network, Bayes classifier, fuzzy logic, and crossbreed calculations can likewise be utilized.

Raj Kumar and Sowrirajan [5] have proposed image-processing-based way to deal with consequent order of the ordinary or infected leaves (Early leaf spot, late leaf spot, alternaria leaf spot), and furthermore give the cure to a similar which would be gainful to novices in cultivating or planting (as these maladies are basic in blossoming plants like rose too). In this approach, they have joined all the mixture highlights of a leaf to prepare the ANN (BPN-FF) and have made utilization of Lloyd's bunching which is more productive than the customary K-implies grouping to section the test pictures.

Dheeb Al Bashish et al. [6] proposed that the RGB images are converted into HSI plane, and then the color features are extracted (by SGDM generation). The texture features are extracted by obtaining Gray-Level Co-occurrence Matrix (GLCM). The information pictures are divided utilizing K-means clustering system, and afterward the sectioned pictures are broke down by a preprepared BPN arranged for discovery and order of plant leaf sicknesses. The author also compares between various models incorporating various components such as HS, H, S, I, and HSI and found that model HS provides the best efficient output among all other models with the efficiency of 92.7%.

Niket et al. [7] stated that the RGB images upon acquisition undergo color space transformation into HSI plane and upon segmentation using K-means clustering the green pixels are masked from the appropriate cluster, and the masked green pixels are removed. Then, the useful segments are obtained. Then, the texture features are extracted using CCM. The classifier used is BPN-FF. But additional color or shape features or both might have improved the efficiency of classification.

Arivazhagan et al. [8] proposed a new method in which the RGB images are converted into HSI plane, and only the hue component is used for further analysis. Then, the green pixels are masked, and the masked green pixels are removed. The useful components are obtained upon segmentation, and only the texture feature is extracted using co-occurrence matrix. Then, the neural network employing SVM classifier is used to detect and classify with the efficiency of 94.74%. Here, only a single feature extraction is employed and the classifier is not that efficient in classifying the disease but effectively detects whether the leaf is diseased or not.

Kaur and Himanshu [9] provided the study of various classifiers, namely, K-Nearest Neighbors (KNN) classifier, Support Vector Machines (SVM) classifier, Backpropagation Neural Network-Feedforward (BPN-FF) classifier, Probabilistic Neural Network (PNN) classifier, and General Regression Neural Network (GRNN) classifier.

Sanjeev et al. [10] provided a new way in which the k-means segmentation is done, followed by feature extraction using GLCM and the classification is done via BPN. Here, only hue component is used and instead of k-means segmentation, other algorithms could have employed to extract lesion more accurately.

Ravichandran and Koteeshwari [11] built android app for the crop prediction which suggests the suitable crop to farmer based on soil parameters, which were entered in application by farmer after soil testing. Here, automatic soil testing is not performed and there is no check on the current crop status.

3 Proposed System

We are proposing a system which is a combination of hardware and software. For obtaining the physical attributes, we use sensors like pH and soil moisture sensors, which are interfaced with controller. Here, we are using raspberry pi controller. Figure 1 shows the overview of proposed system.

Digital camera is used to capture the images from farm and transfer them to computer where image processing is performed in order to detect any diseases on the plant leaves. If starting of disease is detected, then appropriate fertilizer is sprinkled on crop using fertigation pump. As we are having pH and moisture-related data of soil, suggestion can be given to farmer which crop is suitable for that soil. We have gathered the dataset of over 200 high-resolution images of leaves. First, preprocessing of images is performed to remove any noise in image to perform enhancement of image; then, we perform segmentation using the conversion of image from RGB to YC_BC_R plane and feature extraction of where we have considered features such as mean, standard deviation, kurtosis, and skewness for texture features and discrete waveform transformation for the color feature extraction. Classifier utilized is minimum distance classifier which is utilized to arrange obscure picture information to classes which limit the separation between the picture information and the class in multi-include space. The separation is characterized as a record of comparability; so,

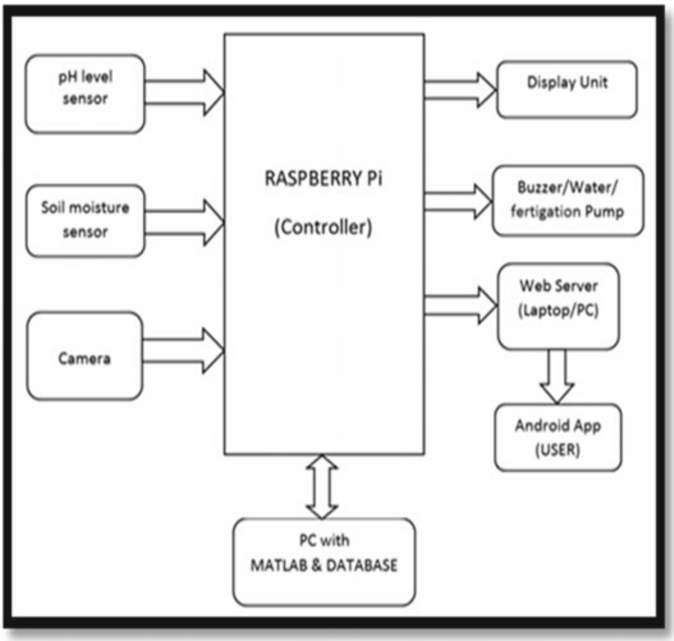


Fig. 1 Block diagram of the system

the base separation is indistinguishable to the greatest closeness with the reference picture, in light of which choice is taken if leaf is having any illness or in solid state.

4 Result Analysis

The result of the proposed system is given below.

After training the dataset, the preprocessing of the input image is done by using the top hat and bottom hat approach. Figure 2 shows the top hat and bottom hat enhancement results. The enhancement is used for enhancing the features of the raw image; after the enhancement, the features of the input image get enhanced and the level of noise gets reduced [12].

After the enhancement, the image segmentation is performed which is used to segment the image into various contours; after the segmentation, the features are extracted. The output of YC_bC_r model is given in Fig. 3.

$$Y = (77/256)R + (150/256)G + (29/256)B \quad (1)$$

$$C_b = -(44/256)R - (87/256)G + (131/256)B + 128 \quad (2)$$

$$C_r = (131/256)R - (110/256)G - (21/256)B + 128 \quad (3)$$

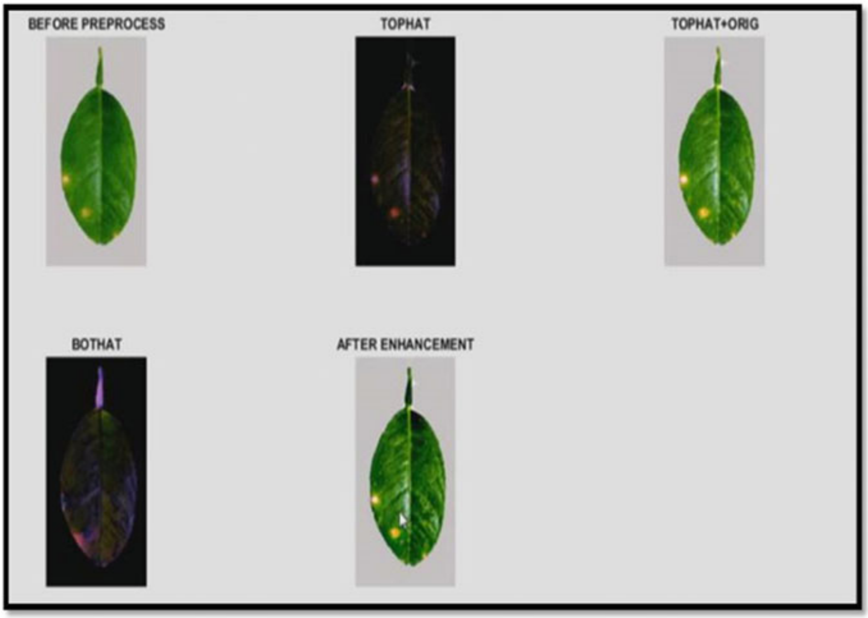


Fig. 2 Output of final enhancement

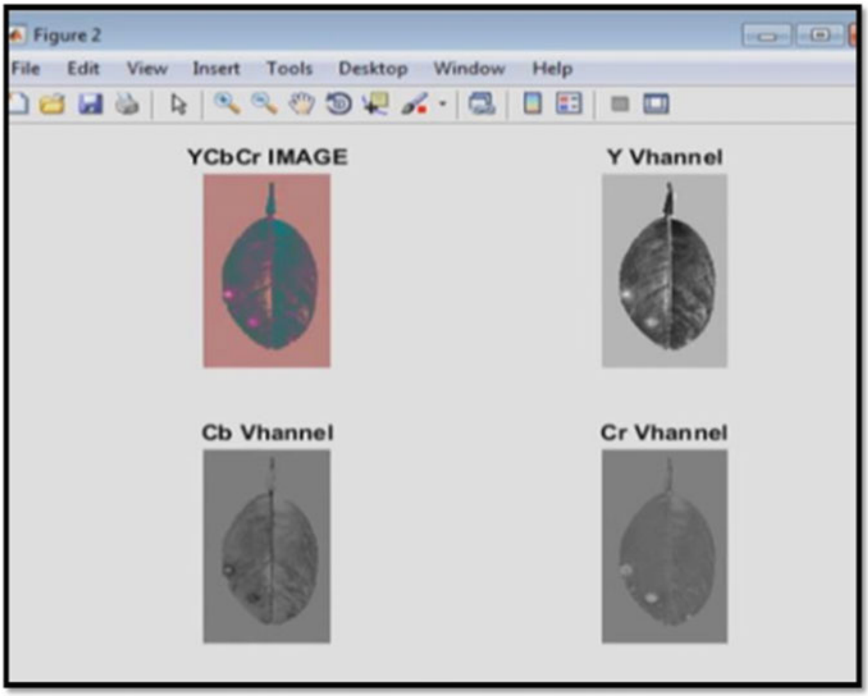


Fig. 3 Output of YC_bC_r model

We have calculated the features of image matrix for example. Mean for a random variable vector A is made up of N scalar observations; the mean is defined as [13].

$$\mu = \frac{1}{N}\sum_{i=1}^{N} A_i \tag{4}$$

The standard deviation is an apportion of how spread out quantities are.

$$y_j = \sigma_j = \sqrt{\frac{\sum_{i=1}^{M}\left|u_{ij} - \mu_j\right|^2}{M-1} \quad 1 \le j \le N} \tag{5}$$

Skewness is a measure of the asymmetry of the information around the example mean. In the event that skewness is negative, the information are spread out more to one side of the mean, i.e., to left than to other side. On the other hand, if skewness is certainly positive, the information are spread out to other side. The skewness of the ordinary conveyance (or any flawlessly symmetric dispersion) is zero. The skewness of a distribution is defined below:

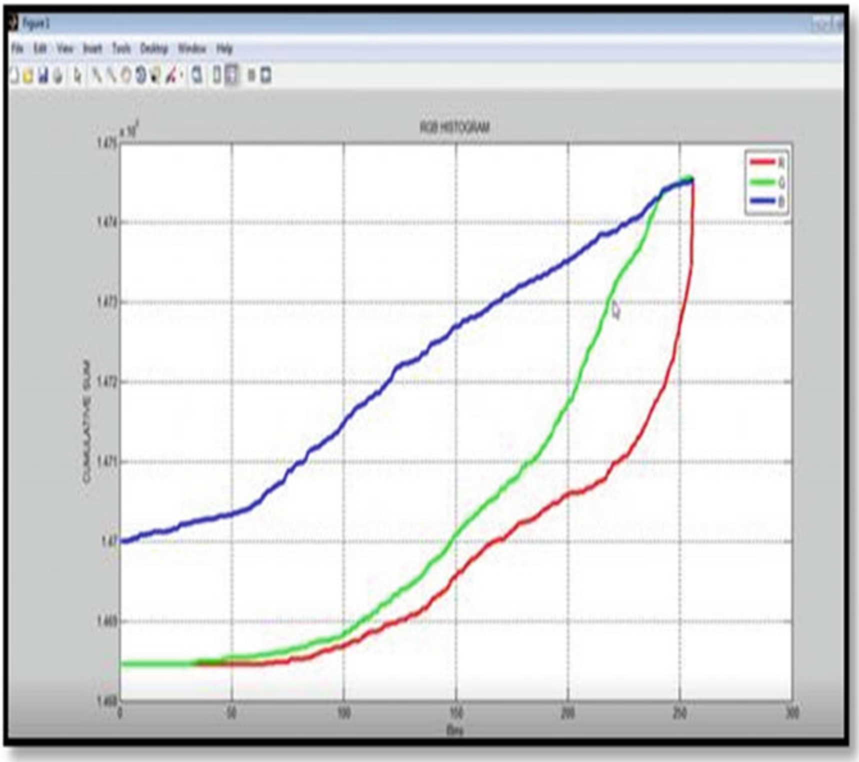


Fig. 4 RGB histogram

$$s = \frac{E(x-\mu)^3}{\sigma^3} \tag{6}$$

Kurtosis is a measure of how an outlier-prone distribution is. The kurtosis of the normal distribution is 3. Distributions that are more outlier-prone than the normal distribution have kurtosis greater than 3; distributions that are less outlier-prone have kurtosis less than 3. The kurtosis of a distribution is defined as (Fig. 4).

$$k = \frac{E(x-\mu)^4}{\sigma^4} \tag{7}$$

After the segmentation, the feature extraction is done; here, we are calculating the color feature and color moments using the RGB to HSV conversion; the output of the color feature extraction is shown in Fig. 5. The C_r channel image is multiplied with complete back image for masking and we can see only diseased part of leaf; however, this image is without color; therefore, we again multiply it with enhanced image pixel by pixel and we got below color segmented image (Fig. 6).

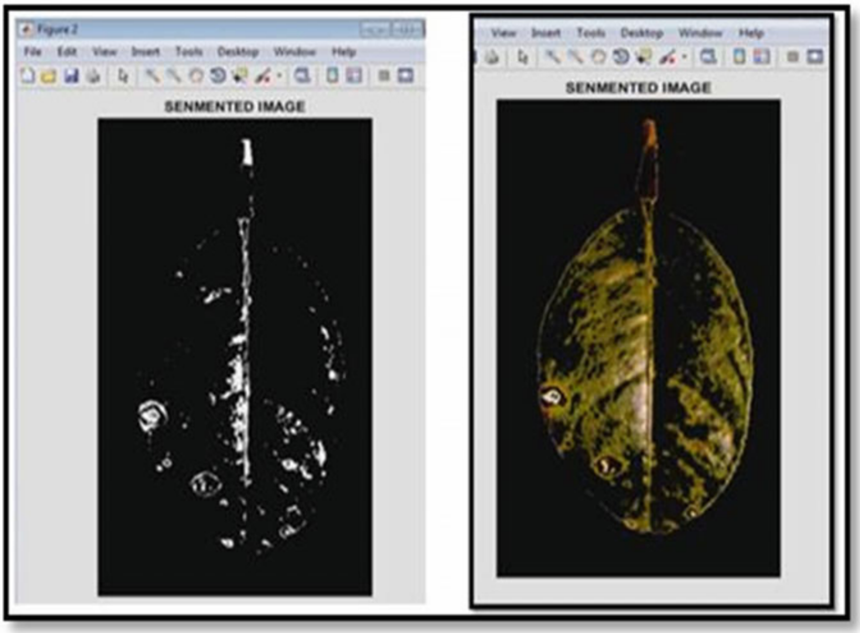


Fig. 5 Color segmented image

After color feature extraction, we move on to extract the texture feature. For this purpose, we make use of the discrete wavelet transform. The discrete wavelet changes state to wavelet changes that the wavelets are disconnectedly evaluated. A change which restrains a capacity both in space and scaling has some important properties contrasted with the Fourier change. The change is fixated on a wavelet lattice, which can be figured more rapidly than the practically equivalent to Fourier framework. Most outstandingly, the DWT [14] is utilized for flag coding, where the advantages of the change are abused to connote a discrete flag in an additional repetitive shape, regularly as a preconditioning for information pressure. The discrete wavelet change has a tremendous amount of uses in science, software engineering, arithmetic, and designing. The mean and standard deviation of the dispersion of the wavelet change coefficients are utilized to build the element vector.

5 Conclusion

To keep up efficiency development in a manageable way, there is a need to move from input concentrated to innovation serious and expertise escalated horticulture. This diminishes the product misfortune by appropriately instructing the ranchers with the harvests points of interest and their prerequisites. The MATLAB ANN Toolbox

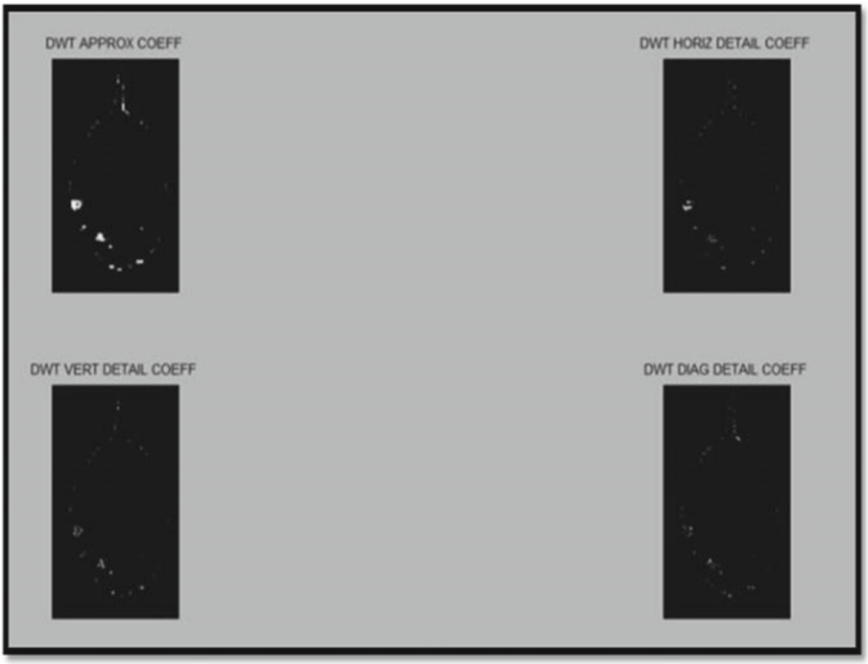


Fig. 6 Output of feature extraction

is the product space in which the expectation framework is constructed and after that imported to Android stage. Most imperative favorable position of utilizing this technique is that the plant maladies can be recognized at the beginning time or the underlying stage.

References

1. Vibhute A, Bodhe SK (2012) Applications of image processing in agriculture: a survey‖. Int J Comput Appl 52(2):34–40
2. Aware AA, Joshi K (2016) Crop and weed detection based on texture and size features and automatic spraying of herbicides. Int J Adv Res Comput Sci Softw Eng 6(1)
3. Aware AA, Joshi K (2015) Wavelet based crop detection and automatic spraying of herbicides. Int J Innovations Adv Comput Sci 4(2)
4. Singh V, Misra AK (2017) Detection of plant leaf diseases using image segmentation and soft computing techniques. Inf Proc Agric 4:41–49
5. Raj Kumar S, Sowrirajan S (2016) Automatic leaf disease detection and classification using hybrid features and supervised classifier. Int J Adv Res Electr Electron Instrum Eng 5(6) https://doi.org/10.15662/ijareeie.2016.0506008
6. Al Bashish D, Braik M, Sulieman B (2010) A framework for detection and classification of plant leaf and stem diseases. In: IEEE international conference on signal and image processing

7. Niket A, Jadhav B, Smeeta N (2014) Detection and classification of plant diseases by image processing. Int J Innovative Sci Eng Technol 1(2)
8. Arivazhagan S, NewlinShebiah R, Ananthi S, Vishnu Varthini S (2013) Detection of unhealthy region of plant leaves and classification of plant leaf diseases using texture features. CIGR J 15
9. Kaur G, Himanshu M (2012) Classification of biological species based on leaf architecture. Int J Comput Sci Inf Technol Secur (IJCSITS) 2(2)
10. Sanjeev S, Vijay S, Nargund VB (2013) Pallavi–Diagnosis and classification of grape leaf diseases using neural network. In: IEEE-2013 4th ICCCNT, July 4–6, 2013
11. Ravichandran G, Koteeshwari RS (2016) Agricultural crop predictor and Advisor using ANN for Smartphones. In: IEEE international conference on emerging trends in engineering, technology and science (ICETETS)
12. Malik S, Lone T (2014) Comparative study of digital image enhancement approaches. In: International conference on computer communication and informatics. http://dx.doi.org/10.1109/iccci.2014.6921749
13. Antoni M, Barlaud M, Daubechies I (1992) Image coding using Wavelet transform. IEEE Trans Image Process 1(2)
14. Kiani1 S, Azimifar Z, Kamgar S (2010) Wavelet based crop detection and classification. In: ICEE 2010, May 11–13, IEEE 2010

Cross Domain Recommendation System Using Ontology and Sequential Pattern Mining

S. Udayambihai and V. Uma

Abstract Recommendation system is very helpful to filter the information according to the user interest and provide user personalized suggestion. Recommendation system is emerging now-a-days in many social networks like Facebook, Twitter, e-commerce etc. Cross domain recommendation system is one of the method to develop the recommendation where we can gather the knowledge from different domains and recommend most similar items related to the user search term. In this work, we try to extend cross domain recommendation by finding semantic similarity of items in Ontology, applying Collaborative Filtering and recommending user preferred items using PrefixSpan algorithm. The similarity between items can be achieved through modified Wpath method. Finally, we can recommend the most preferred items and evaluate using performance measures like F-score.

1 Introduction

Recommendation system is a subclass of information filtering system which is used to check the user 'preference' or 'rating' given to an item. In the e-commerce world, it is used in many fields like movies, music, news, books and research articles. Recommendation system can be implemented using two methods namely Collaborative Filtering (CF) or Content Based Filtering (CBF). Collaborative Filtering (CF) is used to recommend similar items preferred by other users having similar taste. Content Based Filtering (CBF) is alien to CF and is used to match the content according to the user characteristics [1].

Recommendation system is used to retrieve the user preferred information on the internet. By using this, we can increase the average order value and can easily reduce the traffic in services and improve the delivery of relevant content to the user.

S. Udayambihai (✉) · V. Uma (✉)
Department of Computer Science, Pondicherry University, Pondicherry, India
e-mail: itsudaya09@gmail.com

V. Uma
e-mail: umabskr@gmail.com

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_173

Cross domain recommendation systems have the capability to access information belonging to one or more domains [1]. Recommendations can be improved by exploiting the knowledge from source domains and enhancing the recommendations in a target domains. By applying this we can accomplish better accuracy and overcome data sparsity problem [3].

Ontology is used for domain knowledge representation. Ontology represents concepts and relationships between facts in different domains. Ontology is used to describe the individuals, classes, attributes and relations. Ontology can also be represented in the form of Knowledge Graphs (KG) [8].

Semantic similarity is a measure that indicates the similarity between two different entities, sentences, words or documents. Semantic similarity can be applied for topology of ontologies [8]. There are two major approaches for calculating the similarity between words: Pairwise and Group-wise. In pairwise approaches, similarity between two set of words can be calculated by combining the similarity of their words. In group-wise approaches, there are three representation methods viz. set, graph or vector. By using Collaborative Filtering (CF) in recommendation systems, helps to reduce the overloaded information between domains. Collaborative Filtering (CF) when combined with semantic similarity, more accuracy can be achieved with the better recommendations [9].

In this work, ontology is used for representation of knowledge about 2 domains (Music and Books). Semantic similarity is used to measure the similarity between concepts belonging to multiple domains and retrieve the most similar items [11] for recommendation. The proposed cross domain recommendation system, will recommend the items based on semantic similarity measurement [1].

Thus, this proposed work can recommend similar items in E-commerce by involving the user-item-ratings. Section 2 discusses the literature survey. Section 3 explains the system architecture and modules involved. Section 4 briefly explains the experimental setup and Sect. 5 concludes the paper.

2 Literature Survey

Various research works have been carried out related to cross domain recommendation system. Table 1 gives a comparative analysis of some of the related works.

In [1], items are recommended by using User-item rating matrix for Movie-Book domain and by using WordNet based ontology for representing the knowledge. Items are recommended by clustering to form a latent space between different domains. Cosine similarity measure is applied to retrieve the most similar items followed by Collaborative Filtering. But the accuracy is very low due to data sparsity problem. In [3] items are recommended by constructing user-tag-item matrix for Movie—Book domain. Similarity is calculated based on ratio of frequency of occurrence of words and total number of words occurred by applying the k-means clustering and Semantic fuse algorithm the most similar items are recommended. In [8], Semantic Similarity

Table 1 Comparative study of various recommendation approaches

Title and year of publication	Technique used	Evaluation	Knowledge represenation for semantic similarity	Cross domain
[1] Semantic clustering based cross domain recommendation—2014	Clustering to form a common latent space and collaborative filtering	Cosine similarity is used to measure the closeness between source and target domains	WordNet ontology	Movie-book domain for User-item-rating
[2] Personalizing health and food advices by semantic Enrichment of multilingual cross-domain questions—2015	String matching algorithm	Semantic techniques such as OWLIM and sesame RDF	User profile ontology	Health and food domain
[3] Cross domain recommendation using semantic similarity and tensor decomposition—2016	Tensor decomposition, k-means clustering and semantic fuse algorithm	Semantic similarity based on frequency of occurrence of words divided by total number of words occurred	Semantic matrix	User-tag-item for Movie—book domain
[4] Exploiting trust and usage context for cross-domain recommendation—2016	Item based collaborative filtering using association matrix	Computing the cosine similarity of item vectors as the similarity for each item pair	Rating matrix user-item-matrix into item-item matrix association	DVDs, Beauty, Books, Travel and CiaoCafe
[5] Multi-domain collaborative recommendation with feature selection—2016	Feature selection and collaborative filtering	Cosine distance is used to measure the similarity between users using rating distribution and normalization of the ratings	Construction of user-item rating matrix	Two multi-domain product review datasets are used, i.e., Ciao and Epinions
[6] Enhancing cross domain recommendation with domain dependent tags—2016	Cross domain collaborative filtering in matrix factorization	Similarity is calculated using Pearson coefficient	Apply spectral clustering to align domain dependent tags as user-tag and item- tag	User-tags-rating for movie and book

(continued)

Table 1 (continued)

Title and year of publication	Technique used	Evaluation	Knowledge represenation for semantic similarity	Cross domain
[7] Clustering based transfer learning in cross domain recommender system—2016	k-means clustering algorithm and hierarchical agglomerative algorithm	Similarity is measured using cosine of the angle between two vectors	User-item-rating matrix	Tourist and location domain
[8] Computing semantic similarity of concepts in knowledge graphs	Using aspect based sentiment analysis category classification	Similarity is calculated using Wpath semantic similarity. Wpath is measured using shortest path length and contribution of the Least common subsumers of information content	Using knowledge graph to represent ontology	Restaurant domain
[9] A user based cross domain collaborative filtering algorithm based on a linear decomposition model—2017	Collaborative filtering for user-item rating matrix	Semantic similarity can be measured using Pearson correlation with items and ratings on item	Compute the similarity using rich ratings	Book and music domain
[10] Merged ontology and SVM—based information extraction and recommendation system for social robots—2017	Support vector machine (SVM) system	Positive and negative polarities	Merged ontology is the combination of Medical ontology, City ontology and Hotel ontology	Diabetic drugs, hotel and tourism domain
In proposed system, cross domain recommendation system using ontology and sequential pattern mining	Collaborative filtering and prefix span algorithm	Semantic similarity can be measured using modified Wpath method	Two different domain knowledge in ontology	Database for movie from (Movie Lens) and book from (Goodreads dataset)

between different items in the restaurant domain is represented by Knowledge Graph (i.e.) Ontology. Semantic Similarity is calculated using Wpath method,

$$\mathrm{Sem}_{\mathrm{wpath}}\left(c_i, c_j\right) = \frac{1}{1 + \mathrm{length}\left(c_i, c_j\right) * k^{IC(c_{\mathrm{lcs}})}} \quad (1)$$

By using this method we can get more accuracy comparing with the other existing methods. This method is also evaluated using datasets of Aspect Based Sentiment Analysis and also shown the best performance in terms of accuracy and F-score. In [9], items are recommended by constructing user-item rating matrix. Pearson correlation is applied for calculating semantic similarity for Movie-Book domains. Collaborative Filtering method is also applied for retrieving the most similar items. This method can overcome the cold start problem. This method can also be improved by extending the similarity range from local to total similarities.

In this proposed system, item recommendation is done in e-commerce by using Ontology and Sequential Pattern Mining. Cross domain recommendation is provided for two different domains viz. Movie and Book. Ontology is used for knowledge representation. Modified Wpath method is used to calculate the semantic similarity and to improve the results Collaborative Filtering is used in item-rating matrix. Prefix Span algorithm is used to retrieve the Sequential patterns of most preferred items by the regular users. Hence this method is expected to increase the accuracy compared to the existing methods and reduce the data sparsity problem.

3 System Architecture and Modules Involved

3.1 Architecture

The architecture of the proposed cross domain recommendation system is shown in Fig. 1.

3.2 Modules Involved

Module 1: Calculate Semantic Similarity

Semantic similarity is used to measure the similarity between the input like movie name (i.e.) auxiliary domain and the other items in the target domain like books with the help of generated ontology. It also helps in retrieving the most similar items in the target domain.

Here, semantic similarity measure is calculated by using modified Wpath method wherein, length between the two items and height of the Least Common Subsumer (LCS) are considered. It helps to identify the genre and by using this we can easily

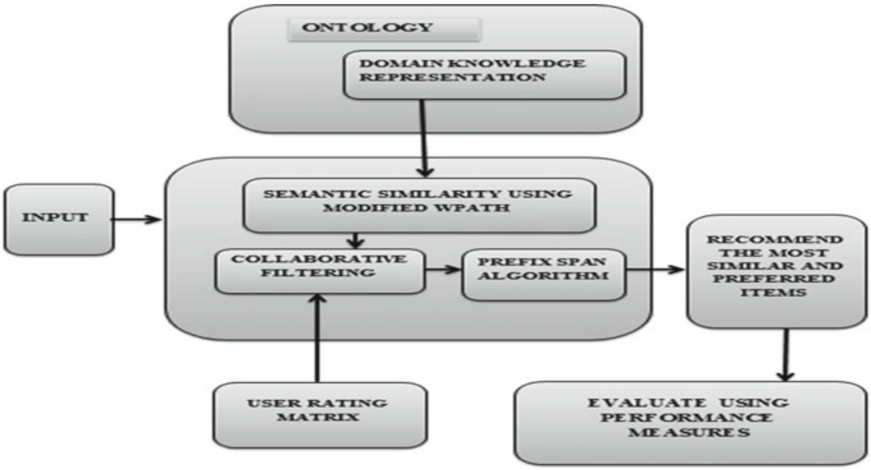


Fig. 1 Proposed system architecture

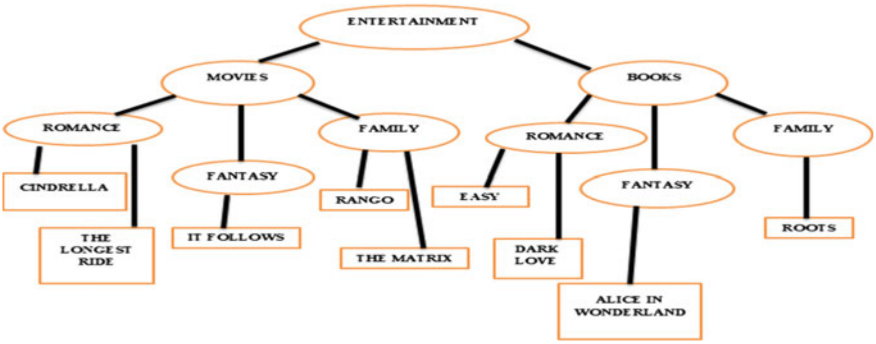


Fig. 2 Fragment of movie and book domain dataset represented as ontology

generate the user rating matrix. Figure 2 represents the movie and book domain in the form of ontology. If Cinderella is given as search term from auxiliary domain then the semantic similarity will be calculated between Cinderella and Easy, Dark love from target domain as both books and movies are from the same genre Romance.

Semantic similarity can be calculated using Eq. (1) by modifying the Wpath method [8].

$$\text{modified Wpath }(c, c_i) = \frac{1}{1 + \text{length}(c, c_i) * \text{height(lcs)}} \tag{1}$$

where c is term given by the user in auxiliary domain movies and c_i is the item of the same genre in the target domain books and length (c, c_i) is the shortest path between

c and c_i and height (lcs) is the maximum height from the least common subsume (lcs).

Module 2: Collaborative Filtering
Collaborative Filtering is used to recommend the items. In item based Collaborative Filtering, similarities are calculated from item-rating matrix. First, write the user-item rating data in the form of matrix. Then create the item-item similarity matrix. The item-item similarity can be calculated using cosine similarity measure.

In this paper, first we have to apply the semantic similarity measure in the Ontology and we have to calculate most similar books that is related to given movie and recommend the item-rating matrix according to the most similar rates.

By using this item-rating matrix apply item based collaborative filtering to find the item-item similarity matrix. To find item-item similarity we can apply cosine similarity measure. The cosine similarity between two item vectors (A, B) can be calculated using Eq. (2).

$$\cos(A, B) = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2}\sqrt{\sum_{i=1}^{n} B_i^2}} \tag{2}$$

where A and B are the ratings given by 'n' number of users for 'm' items. By using this item-item rating matrix we can easily retrieve the most similar items.

Module 3: Apply Prefix Span Algorithm
After applying Collaborative Filtering we go for a new pattern growth method for mining the sequential patterns called Prefix Span algorithm. By applying this algorithm, we can recommend the most similar and also most preferred items of the user.

4 Experimental Setup

Dataset
We have used the freely available dataset for Movie domain from Movie lens dataset which contains 100,000 ratings of 9000 movies given by 700 users. The details <User_id; Movie_id; Movie name; rating; genre> are obtained. We have also used the Book domain dataset from Goodreads dataset which contains 411,156 ratings applied to 7000 books by 600 users. From this available dataset we have taken <User_id; Book_id; Book name; rating; genre>.

In Fig. 3, we are comparing No. of Items in Movies, No. of Items in Books, No. of users for Movies and No. of users for Books using the freely available dataset.

In Fig. 4, we are illustrating the user ratings between Movie and Book domain.

The system is to be implemented and evaluated using F-Score.

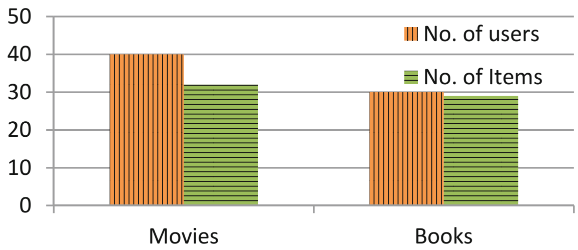


Fig. 3 Illustration between no. of items and users in movie and book domains

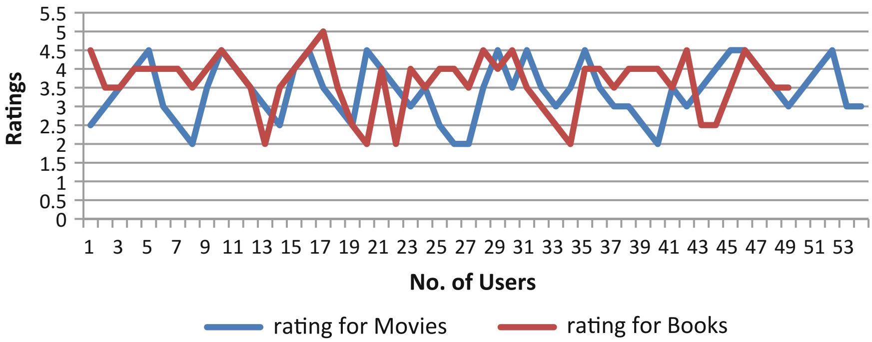


Fig. 4 Illustration of user rating for movies and books

5 Conclusion

By this method we can recommend the most similar items from different domains by combining Ontology with Collaborative Filtering and with the help of user-item-ratings. Modified Wpath method helps in finding the semantic similarity. Prefix Span algorithm is used to retrieve the most preferred items. Hence, Cross Domain Recommendation System using Ontology and Sequential Pattern Mining is expected to achieve more accuracy. Implementing the proposed system and evaluating using the performance measures is left as future work.

References

1. Kumar A, Kumar N, Hussain M, Chaudhury S, Agarwal S (2015) Semantic clustering-based cross-domain recommendation. In: IEEE SSCI 2014–2014 IEEE symposium series on computational intelligence—CIDM 2014: 2014 IEEE symposium on computational intelligence and data mining, proceedings, pp 137–141. http://doi.org/10.1109/CIDM.2014.7008659
2. Al-Nazer A, Helmy T (2015) Personalizing health and food advices by semantic enrichment of multilingual cross-domain questions. In: 2015 IEEE 8th GCC conference and exhibition, GCCCE 2015, pp 1–4. http://doi.org/10.1109/IEEEGCC.2015.7060095

3. Kumar V, Shrivastva KMP, Singh S (2016) Cross domain recommendation using semantic similarity and tensor decomposition. Procedia Comput Sci 85(Cms):317–324. http://doi.org/10.1016/j.procs.2016.05.239
4. Xu Z, Zhang F, Wang W, Liu H, Kong X (2016) Exploiting trust and usage context for cross-domain recommendation. IEEE Access 4:2398–2407. https://doi.org/10.1109/ACCESS.2016.2566658
5. Liu L, Cui J, Song W, Wang H (2017) Multi-domain collaborative recommendation with feature selection. China Commun 14(8):137–148. https://doi.org/10.1109/CC.2017.8014374
6. Hao P, Zhang G, Lu J (2016) Enhancing cross domain recommendation with domain dependent tags. In: 2016 IEEE international conference on fuzzy systems, Fuzz-IEEE 2016, pp 1266–1273. http://doi.org/10.1109/FUZZ-IEEE.2016.7737834
7. Thendral SE, Valliyammai C (2017) Clustering based transfer learning in cross domain recommender system. In: 2016 8th international conference on advanced computing, ICoAC 2016, pp 51–54. http://doi.org/10.1109/ICoAC.2017.7951744
8. Zhu G, Iglesias CA (2017) Computing semantic similarity of concepts in knowledge graphs. IEEE Trans Knowl Data Eng 29(1):72–85. https://doi.org/10.1109/TKDE.2016.2610428
9. Yu XU, Jiang F, Du J, Gong D (2017) A user-based cross domain collaborative filtering algorithm based on a linear decomposition model. IEEE Access 5. http://ieeexplore.ieee.org/abstract/document/8113474/
10. Ali F, Kwak D, Khan P, Ei-Sappagh SHA, Islam SMR, Park D, Kwak KS (2017) Merged ontology and SVM-based information extraction and recommendation system for social robots. IEEE Access 5:12364–12379. https://doi.org/10.1109/ACCESS.2017.2718038
11. Zhang Q, Haglin D (2016) Semantic similarity between Ontologies at different scales. IEEE/CAA J Automatica Sin 3(2):132–140. http://ieeexplore.ieee.org/abstract/document/7451100/

Identifying the Risk Factors for Diabetic Retinopathy Using Decision Tree

Preecy Poulose and S. Saritha

Abstract The role of data mining in healthcare industry is to improve the health systems and to use these data analytics to identify the inefficiencies and to improve care. Accuracy is important when it comes to patient care, and handling this huge amount of data improves the quality of the healthcare system. Diabetic retinopathy is a disease mainly occurring in patients with high sugar level in their blood. The situation occurs when sugar levels in blood are high, thus causing damage to blood vessels. The blood vessels swell and leak and gradually affect the eye vision. This is not detected in early stage of diabetics, even though it affects the eyesight from the beginning. Decision tree classification helps to detect the problem in the initial stage that helps to find the risk factors that cause this disease. Better treatment is provided to those infected patients. A detailed analysis of the result from the decision tree classifier is also presented in this work, and decisive factors for diabetic retinopathy are concluded herewith.

1 Introduction

Diabetic Retinopathy (DR) is a well-known leading cause of preventable blindness, affecting more than 520 million people across the world. DR can be detected at the initial stages by expert doctors through routine analysis of the eye fundus [1]. DR causes serious issues to the retina, and the blood vessels lead to swell and leak. The diabetic patients with prolonged years have more chance to get affected by DR. The patients with high blood pressure and cholesterol, abnormal protein rate in urine, and less oxygen level in retina are more prone to the disease. The rest of the symptoms are scotoma [2] and blurred vision. These symptoms are not identified at the earlier stage. So there is more chance for the patient to get completely blind when not properly treated at the right time. No symptoms are shown at the beginning stage. Later due

P. Poulose (✉) · S. Saritha
Rajagiri School of Engineering and Technology, Kakkanad, Kochi 682039, Kerala, India
e-mail: preecyp@gmail.com

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_174

to prolonged years, the patient feels pain on eyes, swelling on the retina, and blurred vision.

The retina is a well-known source of naturally occurring molecule that enables the early identification of several human diseases, disorders such as high rate in blood pressure, heart diseases, and DR [3]. DR is caused by the damage that has been occurred in the retina due to diabetes. Before 126.6 million people have been diagnosed with DR, and now showing a growth of up to 190 million people by 2030 as per [http://en.wikipedia.org/wiki/Diabetic_Retinopathy], accessed on March 30, 2018.

The algorithms used are random forest and decision tree classifier. This algorithm is used to find out whether the patient will be affected by DR. A simpler way to find the risk factors is using the random forest algorithm. DT is used to identify the decisive factors that affect the DR [4].

The paper is organized into the following sections: Sect. 1 gives the entire introduction of work that is being proposed. Section 2 briefs about literature review of current work. Section 3 describes the proposed system design and architectures. The results and discussions are presented in Sect. 4 and concluded in Sect. 5.

2 Literature Survey

Previously published works were used as a guideline to carry out the survey. Several classifiers have been used for the prediction of risk factors affecting DR. They are classified on the basis of accuracy. The important concepts are discussed below.

In the research work, the non-dilated retina images are taken as an input for diagnosing the diabetic retinopathy [5]. Attribute selection plays a major role in decision tree classifier for classifying the input image data. The input data are then preprocessed and based on the result the images are classified into infected and non-infected images. If the image is infected, it is then given for further processing and classified them into severe infected and moderate infected image. The feature extraction or attribute selection is used for both the train data and the test data. By calculating the accuracy, the performance of the GMM classifier is obtained and identified [5].

DR occurs mostly in both middle-aged and aged people. The diagnosing of the disease at the early stage and providing proper treatment and awareness at the right time can slow down the progression of disease and prevents from complete blindness or vision loss [6]. An automated detection method can be used by the non-experts to detect which patient needs to be referred for an ophthalmologist. It reduces the burden on ophthalmologist, in order to improve the efficiency of diabetic screening program and reduction of cost. PNN classifier [6] is successfully used to detect the exudates in retinal images.

3 Design of Proposed System

The system architecture mainly consists of four modules including data, data processing, decision tree classifier, and lastly classifier evaluation. Here in the project, the input data are taken as the raw data regarding the medical report and symptoms of diabetic patient. The raw data are then preprocessed and then given to the decision tree classifier, and finally the results are evaluated. The two-third data are given as the training data, and one-third data are given as the test data. Decision tree classifier is implemented on both the data, and finally the result is being compared and evaluated (Fig. 1).

3.1 Data Module

The input data are taken from the medical records provided in the UCI repository. These data include the features and attributes that cause the DR. The data are provided as a text format.

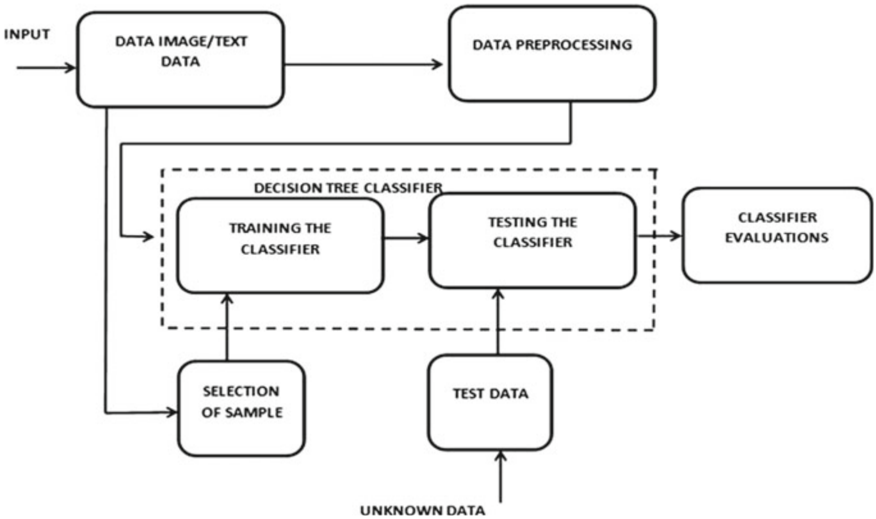


Fig. 1 Overall system architecture of the proposed design

3.2 Data Preprocessing

The raw data are preprocessed to finally give to the classifier. The missing data are filled with the real data, and incorrect data are corrected with real attributes. Some tuples are ignored which have incorrect values.

3.3 Classifier Module

The two-third of data is given for training and one-third is given for testing. Decision tree classifier is used with class labels yes/no.

Train data
The two-third of data from the dataset is taken as an input or train data. It is then given to the decision tree classifier. The algorithm regarding the decision tree will work on these data, and finally it will train the dataset. While training, the last column of attributes is not taken into concern, i.e., the status of the patient, whether the patient will be affected by the diabetic retinopathy or not.

Test data
The one-third set of dataset is given as the test data. And it will then compare the actual result and the predicted result. With this, the final accuracy is being calculated.

Decision tree classifier
The decision tree classifier uses an ID3 algorithm to find out the decisive factors that mainly affect the diabetic retinopathy. The factors are determined by calculating the entropy of each attribute from the dataset. The attribute with the highest entropy is then selected as the decisive factor, and then further classification is provided on this attribute to find the succeeding nodes and finally, the algorithm terminates when the nodes get equal to the class labels.

3.4 Classifier Evaluation Module

The classifier is evaluated based on three categories and they are accuracy, precision, and recall. Two-third of data is given for training and one-third is given for testing. If results are not satisfactory, then training samples are improved.

Random Forest classifier
The advantage of using the algorithm is its problem-solving capability, mainly in regression problems.

Decision tree Classifier
One big advantage of using decision tree classifier is its ability to solve problem, decisions, and outcomes of decision.

4 Results

The dataset includes the severe symptoms that occur in a diabetic patient. These data are in a text format, and the attributes included in the dataset are patient id, age, gender, prolonged year, glucose level before fasting, glucose level after fasting, HbA1c, blood pressure, protein in urine, triglycerides, scotoma, pain on eyes, macular oedema, and oxygen level on retina. These data are then undergone a preprocessing stage that added up the missing values and ignored the incorrect values or tuples. The implementation is done in Python 2.7.9.

4.1 Sample Decision Tree Implementation

See Fig. 2.

4.2 Analysis of Decision Tree

On analysis of the decision tree, the decisive factors that help in identifying diabetic retinopathy are (a) scotoma, (b) pain on eyes, (c) triglycerides, and (d) prolonged

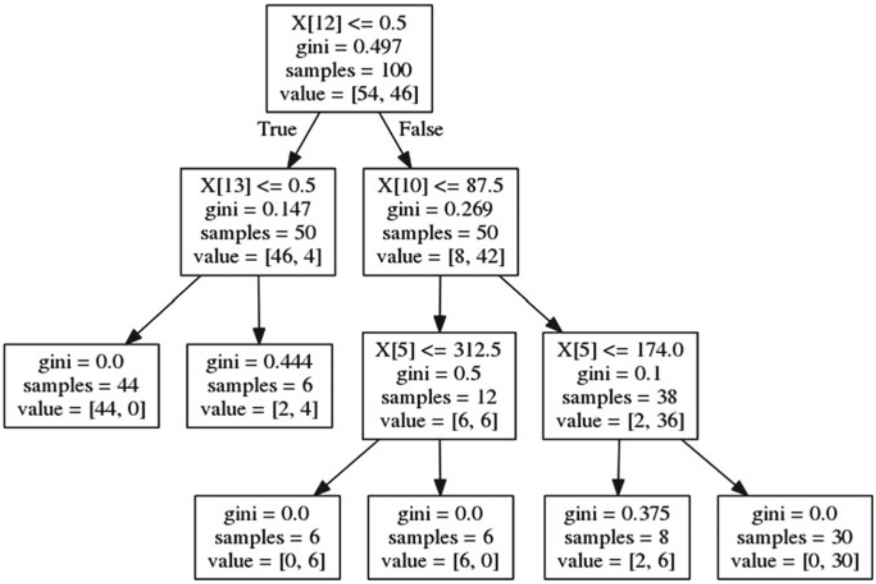


Fig. 2 Decision tree

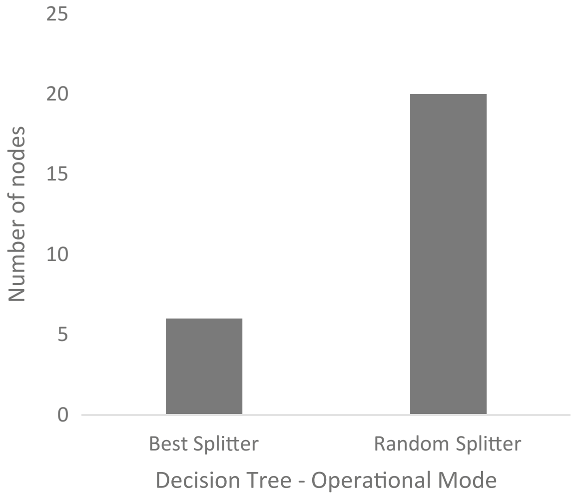


Fig. 3 Operational mode of decision tree

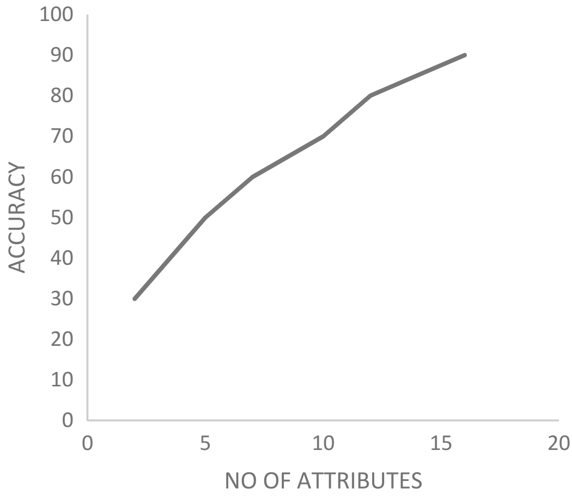


Fig. 4 Accuracy plot for different number of attributes

year of diabetes in the order of importance. The factors have been concurred with the aid of a domain expert.

The graphical representation of operational mode of decision tree is given in Fig. 3. As per the splitter being used in the algorithm, the nodes in the decision tree can vary. Best splitter is used to find the best decisive factors that affect DR. While using this best splitter, the number of nodes is reduced in the decision tree.

The accuracy of decision tree obtained is dependent on number of attributes in the input dataset. As the number of attributes increases, the accuracy associated also increases. A graphical representation of the same is presented in Fig. 4 (Table 1).

Table 1 Relation between attributes and accuracy

Number of attributes	Accuracy (%)
6	56.33
8	77.32
10	88.21
12	92
14	93.21
16	97.57

5 Conclusion

This work presented a data mining method to find the occurrence of diabetic retinopathy in diabetic-affected patients in the early stage itself. The factors causing diabetic retinopathy is identified in this work with the help of decision tree classifier. The decisive factors identified from this work were verified with a subject expert and was appreciated as commendable. A detailed analysis of the decision tree obtained is also done in the experiments. As a future scope, the input data of the design can be changed to retina image and appropriate image mining techniques can be applied to obtain better results.

References

1. Zhang Y, Tsai C-W, Hassan MM, Alamri A (2017) Health-CPS: healthcare cyber-physical system assisted by cloud and big data. IEEE Syst J 11(1)
2. Hamsagayathri P, Sampath P (2017) Priority based decision tree classifier for breast cancer. In: 2017 international conference on advanced computing and communication systems (ICACCS-2017), 06–07 Jan 2017, Cancer Detection
3. Senthilnayaki B, Venkatalakshmi K, Kannan A (2013) An intelligent intrusion detection system using genetic based feature selection and modified J48 decision tree classifier. In: 2013 fifth international conference on advanced computing (ICoAC)
4. Sarwinda D, Bustamam A (2017) A complete modelling of local binary pattern for detection of diabetic retinopathy. In: 2017 1st international conference on informatics and computational sciences (ICICoS)
5. Dhanasekaran R, Mahendran G, Murugeswari S, Fargana SM (2016) Investigation of diabetic retinopathy using GMM classifier. In: 2016 international conference on advanced communication control and computing technologies (ICACCCT)
6. Mahendran G, Dhanasekaran R (2014) Identification of exudates for diabetic retinopathy based on morphological process and PNN classifier. In: International conference on communication and signal processing, 3–5 Apr 2014

Heavy Vehicle Detection Using Fine-Tuned Deep Learning

Manisha Chate and Vinaya Gohokar

Abstract Heavy vehicles develop technical snag and traffic jam on streets. Accidents between heavy vehicle and road users, for example, pedestrians often result in severe injuries of the weaker street users. The highway safety and traffic jams can be secured with detection of heavy and overloaded vehicles on the highway to facilitate light motor vehicles like cars, scooters. A model for heavy vehicle detection using fine-tuned based on deep learning is proposed to deal with entangled transportation scene. This model comprises two parts, vehicle detection model and vehicle fine-grained detection. This step provides data for the next classification model. Experiments show that vehicle's make and model can be recognized from transportation images effectively by using our method. Experimental results demonstrate that the proposed detection system performs accurately with other simple and complex scenarios in detecting heavy vehicles in comparison with past vehicle detection systems.

1 Introduction

Traffic congestion is another one of the curses of human driving, and in particular of clumsy driving. Heavy vehicles are still using regular roads and will be subject to congestion and space constraints of the roads. Heavy vehicles develop snag while climbing, especially at bridge. If this happens, vehicular movement gets blocked. Roads are pricey and in many cities there is no room to build new roads. Heavy vehicles on roads just end up with waiting in traffic jams. Much congestion, though, is just caused by heavy vehicles in rush hours. Heavy vehicle detection system provides solutions to avoid accidents and traffic jams. Here, we address the problem of heavy vehicle detection and recognition.

M. Chate (✉)
Amity University, Noida, India
e-mail: mmchahande@amity.edu

V. Gohokar
MIT, Pune, India
e-mail: vvgohokar@mitpune.edu.in

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_175

If fine-grained deep learning approach of heavy vehicle detection is applied in transportation and public security, we can acquire more meta information like vehicle make, model, logo, production year, max speed, acceleration, and so on [1]. By acquiring this information dynamically, we can build a large intelligent transportation system that can monitor the whole city's road. Further, we can analyze the heavy vehicles on the road at different times to find the discipline of people's going out, and then we can schedule transportation rules accordingly; these will make cities more smart and intelligent.

The rest of this paper is organized as follows: Sect. 2 outlines some relevant research on vehicle detection and CNN applications; Sect. 3 provides a meticulous description of the proposed system. Implementation details and experimental observations will be provided in Sect. 4, followed by conclusion in Sect. 5.

2 Related Work

Diverse methods for vehicle detection and classification have been invented recently. Sivaraman and Trivedi [2] use active learning to learn from front part and back part vehicle images, and achieve 88.5 and 90.2% precision on front and rear part vehicle detection. Chen et al. [3] use a measurement-based feature (MBF) and intensity pyramid-based HOG (IPHOG) combined feature set for vehicle classification. Kafai and Bhanu propose a classification approach for rear vehicles [4]. They define a set of features that include tail light and plate position information, and then pass it to a hybrid Bayesian grid for classification.

Various vehicle recognition methods for complex environments have been proposed and made some achievements. Generally, the research on vehicle recognition can be classified into two categories: "handcrafted feature and classifier" based methods and deep-learning-based methods. For example, Chen et al. [5] used the symmetric SURFs feature to detect the matching feature pair in the vehicle and classified the vehicles by the sparse classifier, which obtained better classification accuracy. Li et al. [6] applied HOG numbering your section headings.

Histogram of oriented gradients is used to describe the apparent properties of the vehicle and constructed a top-down and-or model combined with SVM to achieve the vehicle detection and recognition. The accuracy of recognition can be improved obviously with the introduction of and-or model. Lin et al. [7] proposed a 3D modeling method based on the deformable parts model (DPM), which is a classic part-based algorithm. This method can generate the vehicle classification model for fine-grained recognition and achieve good results. The latter category of vehicle recognition research is designed to procure classification model from big data using deep learning. The existing research shows that vehicle recognition is capable of improved performance in a very significant manner more effectively by discovering high-level depth features via big data [8]. Compared with the traditional idea of combining handcrafted features and classifier, recognition accuracy has improved significantly. Heikki et al. [9] adopted the end-to-end structure of five-layer convolutional neural network (CNN) network to achieve the vehicle recognition. Wang et al. [10] abandoned the traditional end-to-end deep structure and applied the feature expression of mid-level layers, combined the SVM classifier with the triplet mining method based on the order and shape condition, which achieved the state-of-the-art recognition results which were obtained on the car-196 dataset.

The deep-learning-based methods for vehicle recognition can ward off the process of handcrafted features design and automatically learn the deep features from big data. The high-level features can provide the important data of the category while suppressing the irrelevant background information. The deep-learning-based methods also have the excellent properties of generalization and robustness. Although these methods overcome the impact of the complex environments on vehicle recognition to some extent and increase the recognition accuracy and efficiency, they do not consider the relationship among the data, but only employ the advantage of CNN structure to obtain the classification model brutally. Inspired by previous research aimed at displaying and activating CNNs to carry out special tasks, we focus on exploring the transfer approach that is commonly used in deeper learning applications. The benefit of the transfer of learning is that the number of images needed for training and training time is highly minimized.

3 Transfer Learning

Transfer learning is considered to be the transfer of knowledge from a task learned to a new task in machine learning. In the context of neural networks, it transfers the features learned from a pretrained network to a new problem. The formation of a convoluted neuronal network from the beginning in any case is not usually effective when there is not enough amount of training data [2, 4]. The common practice in deep learning for these cases is to use a network that is trained in a large dataset for a new problem. While the initial layers of the pretrained network can be fixed, these last layers need to be refined to learn the specific features of the new dataset. Learning transfer often involves faster training times than the formation of a new convolutional neuronal network since it is not mandatory to estimate all the parameters of the new network. Transferable learning is useful when you want to form a CNN in your data set, but for various reasons, the dataset may not be suitable for forming a complete neural network (that is, too small). While data growth is a viable option in many cases, transfer learning has also proved effective. Transfer learning refers to the process of assuming a precompressed CNN, which replaces the completely the last convolutional layer and the formation of such layers in the dataset. By freezing convolutional layers weights, deep CNN can extract image features like edges, while fully connected layers can proceed with this information and use it to categorize data relevant to the problem. In practice, when there is a moderate amount of data, transfer learning is done at the last level and the classification layer is completely connected with a 1/100 learning rate of the value used to obtain the model convergence during training in ImageNet.

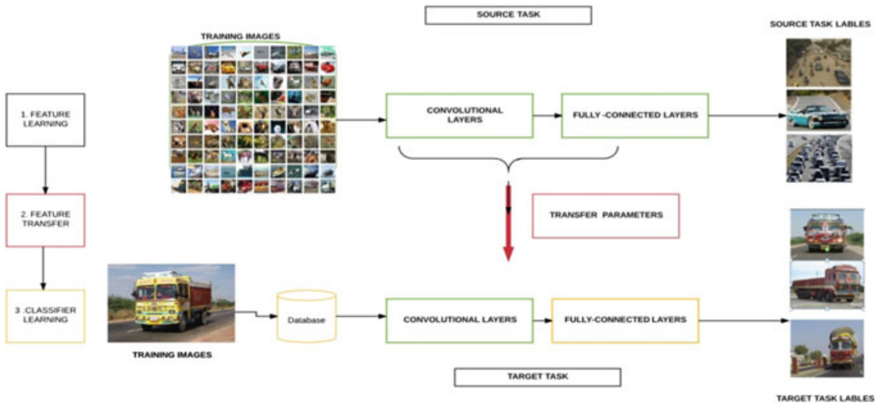


Fig. 1 Network architecture for proposed heavy vehicle detection system

4 Proposed Algorithm

We propose a heavy vehicle detection system based on fine-tuned deep learning. In the initial phase of our system, an input image is applied to pretrained CNN. The image process is from the R-CNN detector. The R-CNN detector only analyzes areas that may contain an object. This greatly minimizes the computational costs incurred while executing a CNN. As soon as transfer learning is used, in the transfer of learning, a network trained in a large collection of images is used as a starting point for solving a new detection activity.
Proposed algorithm can be explained in the following steps:

Step 1: Data collection.
Step 2: Train CNN using CIFAR-10 Data, with initial learning rate of 0.001.
Step 3: Approve CIFAR-10 network training. After the network is trained, it should be validated to ensure that training was successful. First, a swift visualization of the first convolutional layer's filter weights can help determine any immediate issues with training.
Step 4: Input Training Data.
Now that the network is working satisfactorily for the CIFAR-10 classification, the transfer learning methodology can be used to fine-tune the network for heavy vehicle detection.
Step 5: **Load** the ground truth data.
Training data accommodated in a table that contains the image file name and the ROI tags. Each ROI tag is a boundary box around the objects of concern within an image. To train the heavy vehicle detector, only the heavy vehicle image ROI tags are required (Fig. 1).
Step 6: Train R-CNN heavy vehicle detector.
Finally, train the R-CNN object detector using the R-CNN object detector. The training function automatically modifies the original CIFAR-10 network, classified in a

network that can classify images in two classes: heavy vehicle and a generic background class.
Step 7: Visualize the feature map superimposed on the test image.

5 Experiments

First, a CNN was previously trained utilizing the CIFAR-10 data set, which consists of 50,000 training images. So this pretrained CNN is tuned for our heavy vehicle with only 41 on road heavy vehicle images. For testing, we have considered heavy vehicle and non-heavy vehicle dataset.

5.1 Detection Results

We are apt to use image instead of video, because it is easy to measure the accuracy. Our system is able to detect the frontal view of each vehicle image accurately. We performed the detection algorithm on all images in the system, and got 100% accuracy with respect to detection accuracy. Thus, our detection algorithm proved to be effective and fast in the desired system. Once the frontal view of a vehicle was obtained, we fed it into trained deep model as input; the output label is used to recognize the vehicle model. In order to keep the environment constant for comparison, we trained the deep architectures for 1, 10, 20, 50, and 100 epochs. During the training, the base learning rate was set to 0.01.

Figure 2 shows the results of number of epochs with accuracy. Table 2 shows the average accuracy of the CNN cascade method and our approach. Our method shows the best result. Figure 3 shows sample images of heavy vehicle database. Figure 4 shows detected images (Table 1).

Table 2 shows the final result of proposed method for heavy vehicle detection. These images were taken from real street CCTV.

Table 1 The result of AUC of heavy vehicle

Method	AUC (%)
Proposed method	100
Multi-scale CNNs [11]	98.31
Cascaded CNN [12]	96.50

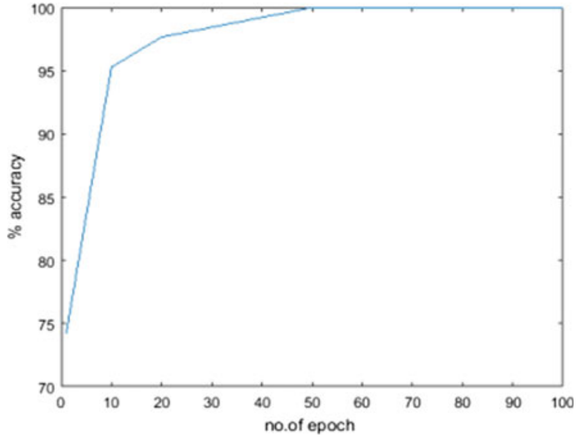


Fig. 2 Loss versus no. of epochs

Fig. 3 Sample images of heavy vehicle database

Fig. 4 Detected images

Table 2 Fine-tuned results on heavy vehicle dataset

Epoch	Time elapsed loss	Mini-batch loss	Mini-batch accuracy (%)	Base learning rate
1	110.65	0.8095	74.22	0.0010
10	1117.10	1587	95.28	0.0010
20	2522.52	0.0479	97.66	0.0010
50	6489.36	0.0074	100.00	0.0010
100	15,698.94	0.0002	100.00	0.0010

6 Conclusion

In this paper, a transfer learning based heavy vehicle detection system proposed. In addition, a reformative CNN structure is presented to enhance recognition accuracy and hugely scale down the computational cost incurred when running a CNN. By introducing transfer learning for proposed system, accuracy rate is significantly improved. Considerable tests have been performed, yielding favorable results.

References

1. Ge Q, Wen C, Duan S (2014) Fire localization based on range-range-range model for limited interior space. IEEE Trans Instrum Meas 63(9):2223–2237
2. Sivaraman S, Trivedi MM (2012) Real-time vehicle detection using parts at intersections. In: 15th International IEEE conference on intelligent transportation systems (ITSC). IEEE, pp 1519–1524
3. Chen Z, Ellis T, Velastin S (2012) Vehicle detection, tracking and classification in urban traffic. In: 15th International IEEE conference on intelligent transportation systems (ITSC). IEEE, pp 951–956
4. Kafai M, Bhanu B (2012) Dynamic bayesian networks for vehicle classification in video. IEEE Trans Ind Inf 8(1):100–109
5. Chen LC, Hsieh JW, Yan Y, Chen DY (2015) Vehicle make and model recognition using sparse representation and symmetrical SURFs. Pattern Recogn 48(6):1979–1998
6. Li B, Wu T, Zhu SC (2014) Integrating context and occlusion for car detection by hierarchical and-or model. In: Computer vision ECCV 2014. Springer International Publishing, pp 652–667
7. Lin YL, Morariu VI, Hsu W, Davis LS (2014) Jointly optimizing 3d model fitting and fine-grained classification. In: Computer vision–ECCV 2014. Springer International Publishing, pp 466–480
8. Xie S, Yang T, Wang X, Lin Y (2015) Hyper-class augmented and regularized deep learning for fine-grained image classification. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 2645–2654
9. Huttunen H, Yancheshmeh FS, Chen K (2016) Car type recognition with deep neural networks. arXiv preprint arXiv:1602.07125
10. Wang Y, Choi J, Morariu VI, Davis LS (2016) Mining discriminative triplets of patches for fine-grained classification. arXiv preprint arXiv:1605.01130
11. Sermanet P, LeCun Y (2011) Traffic sign recognition with multi-scale convolutional networks. In: The 2011 International joint conference on neural networks (IJCNN), July 2011, pp 2809–2813
12. Zang D, Zhang J, Zhang D, Bao M, Cheng J, Tang K (2016) Traffic sign detection based on cascaded convolutional neural networks. In: 2016 17th IEEE/ACIS International Conference on software engineering, artificial intelligence, networking and parallel/distributed computing (SNPD), Shanghai, pp 201–206

Using Contourlet Transform Based RBFN Classifier for Face Detection and Recognition

R. Vinothkanna and T. Vijayakumar

Abstract Face is a highly non-rigid object; in such case, face detection and recognition has become an essential part of biometric systems in the majority of the applications. Numerous applications like robots, tablets, surveillance systems, and cell phones revolve around an efficient face detection and recognition technique in the background for access. Human–computer interaction systems like expression recognition, cognitive state/emotional state, etc. are used. Recognizing with the increased need for security and anticipation of spoofing attacks, almost all techniques have been proposed in the past to successfully detect and recognize the face through a single or combination of facial features, which is a challenging task given the complex nature of the background and the number of facial features involved. Here, the proposed work involves a multi-resolution technique, namely, the Contourlet transform along with linear discriminant analysis for feature detection given to an RBFN classifier for effective classification. It could be clearly seen that the proposed technique outperforms the other conventional techniques by its recognition rate of nearly 99.2%. The observed results indicate a good classification rate in comparison with conventional techniques.

1 Introduction

One of the recent booming technologies in the field of security systems is biometrics which involves identification of a person based on physiological or behavioral traits. By unauthorized access to confidential places, data is being reported throughout the globe. Hence, the need for authenticated entry mechanism has become the need of the hour. Biometric systems offer an effective solution in providing access by

R. Vinothkanna
ECE, Lakshmaiah Education Foundation, Guntur, Andhra Pradesh, India
e-mail: rvinothkannaphd@gmail.com

T. Vijayakumar (✉)
ECE, Guru Nanak Institute of Technology, Hyderabad, India
e-mail: vishal_16278@yahoo.co.in

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_176

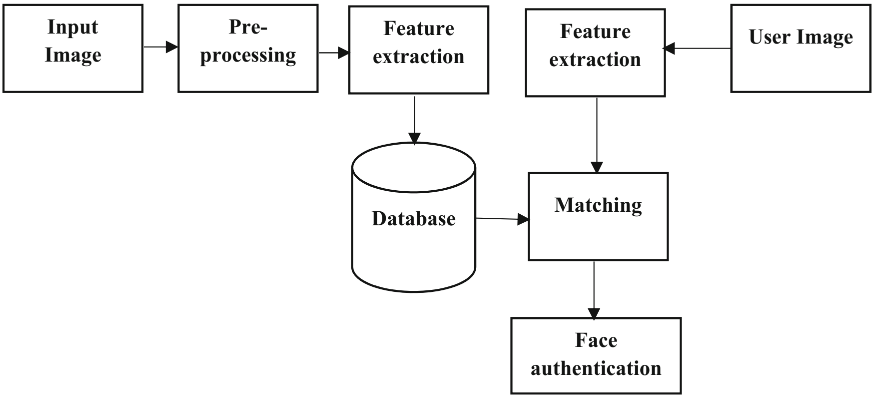


Fig. 1 A simple face authentication system

authenticating the user based on the physical or behavioral characteristics of the particular individual. Physical traits may include facial features [1, 2], fingerprint, palm-print, finger veins, iris, retina, etc., while the behavioral characteristics may include the style of walking referred to as gait of an individual, speech, etc. Since the objective of this paper is to detect an individual based on the facial features, the physical traits are taken into account in this paper. The objective [3] of face detection from face involves several features such as the orientation at the time of capture, the environmental conditions like lighting, etc., the focusing of the image acquisition device, the hair color, skin texture, brow length and arc angle, distance between the brows, the chin protrusion, etc., and the process of face detection involves pre-processing, feature extraction and a classifier as shown in Fig. 1. Any biometric mechanism basically involves two phases namely, the registration and authentication phase. Registration phase refers to the stage where the inputs are fed into the database and trained for the features. Authentication stage refers to the stage where the user is authenticated by comparing the features with the stored features in the database.

The figure illustrates two inputs, namely, the input image and the user image. Input image represents the image used for registration of features into the data base, while the user image denotes the image of user to be authenticated. As shown in Fig. 1, the enrollment phase is where the individuals are registered into the system using the features of their fingerprints. The unique features of each of the individuals are stored in the database. Verification phase takes place before the actual implementation where a known user claims the identity to the access system and verification is done based on the biometric features of a sample taken from the database. In the identification phase, any individual claims identity and access to the system and the biometric system decides on a true or false identity. In the above system, even though the features are unique, their storage in a database always poses a risk of theft leading to spoofing and hacking attacks. Face detection should be performed before recognition system. This is done to extract relevant information for face and facial expression analysis.

2 Related Work

Face detection using Eigen methods is one of the most widely and conventionally used techniques for face detection which comes under the appearance- based approach. They utilize a holistic approach to capture variations in a set of images in combination with principal component analysis (PCA) [4] for feature extraction. The notable demerits are the time consumption for eigen-value computation and also the observed results indicate that they are sensitive to lighting conditions [5]. Euclidean distance has also been used as parameter for feature matching [6] which is observed to detect variations in feature data due to lighting conditions. Recent face detection and classification techniques are mostly based on neural networks [7] due to its feasibility of training a system to capture the complex class of face patterns. Further, they are non-linear and hence the feature extraction step may be more efficient than the principal component analysis. 96.2% accuracy with a classification time less than 0.5 s has been observed [8]. A number of evolutionary variants have been utilized for face recognition like El Bakry's fast neural network [9], Huang's polynomial neural network (PNN), and conventional neural network of Matsugu [10] which has been extensively discussed in [5]. The problem of subject independence and translation, rotation, and scale invariance [11] in facial expression recognition has been addressed in the above variant techniques. With the advent of multi-resolution techniques, there have been more face detection techniques based on wavelet transforms [12]. The wavelets concentrate the energy of image into coefficients and further processing at frequency level causes least visual changes in the human visual perception system. A Gabor wavelet transform [13] addresses the issue of light sensitiveness in previous methods by applying the Gabor wavelets to a local binary pattern (LBP) [14] and experimentations indicated a good acceptance rate in comparison with conventional techniques. Linear discriminant analysis (LDA) techniques have been of late being used for feature extraction and reduction [15, 16]. LDA is a single input analyzer and comes under supervised learning techniques and is a linear technique. It combines the independent feature which leads the largest mean differences between the desired classes.

Hybrid techniques have been developed recently like the works of Sahoolizadeh et al. [17] based on combined Gabor wavelet and ANN feature classifier. The work indicated a drastic reduction in dimensionality, and linear discriminate analysis on down sampled Gabor wavelet faces is found to increase the discriminate ability. A modification was carried out in [18] where Gabor wavelets transform was used in hybrid combination with a feed-forward neural network for finding feature points and extracting feature vectors. The location of feature points contains information about the face in this approach. A correlation- based technique was presented in [19] by estimating areas of candidate of face presence followed by extraction of Gabor wavelets characteristics and neural network classifier. A back-propagation neural network [20] along with a Fourier Gabor filter used color skin to detect face regions and recognition achieved by using BPNN. The conventional support vector machine (SVM) has also been used to determine the match [21].

3 Proposed Work

Face detection is followed by face recognition where the process of face detection involves image preprocessing and feature extraction. A simplified face detection and recognition flow are shown in Fig. 2.

A. PREPROCESSING: Preprocessing is done on the input image to remove noise. In the process of capturing images, distortions including rotation, scaling, shift, and translation may be present in the face images, which make it difficult to locate at the correct position. Pre-processing removes any un-wanted objects (such as, background) from the collected image. It may also segment the face image for feature extraction. Histogram equalization could also be performed to distribute the image intensities throughout the image.

B. FEATURE EXTRACTION AND REDUCTION: Contourlet transform is used for decomposing the input image into sub-bands of differing scales of resolution. A k-level decomposition provides 2^k sub-bands out of which the first half are horizontally oriented and the other half are vertically oriented. The transform is a directional multi-resolution image representation scheme proposed by Do and Vetterli [22] which is effective in representing smooth contours in different directions of an image, thus providing directionality and anisotropy. While the Curvelet transform operates on continuous domain, the Contourlet transform is flexible to be operated in discrete domain. It utilizes a double filter bank; the Laplacian pyramid (LP) is used to detect the point discontinuities of the image and then a directional filter bank (DFB) to link point discontinuities into linear structures. A four-level decomposition of an input image from the ORL data base is shown in Fig. 3.

The Laplacian and directional filter bank are combined to capture the directional information. The texture information is captured in the directional sub-bands and hence the low-frequency sub-bands are ignored. In the proposed work, the LDA is applied to these sub-bands containing the maximum texture features to reduce the dimensionality of the feature vector and capture class-specific features by removing redundant elements. For each directional sub-band, all instances of the same person's face are grouped into one class and the faces of different subjects into different classes for all subjects in the training set, followed by labeling of the instances. Accordingly, within-class and between-class scatter matrices are designed as follows:

$$S_{\mathrm{WC}} = \sum_{c} \sum_{ci} (\mu_c - \overline{x_i})(\mu_c - \overline{x_i})^{\mathrm{T}} \tag{1}$$

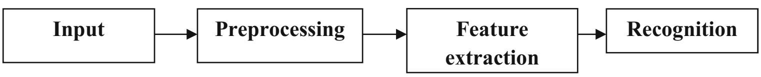


Fig. 2 A simplified face detection and recognition system

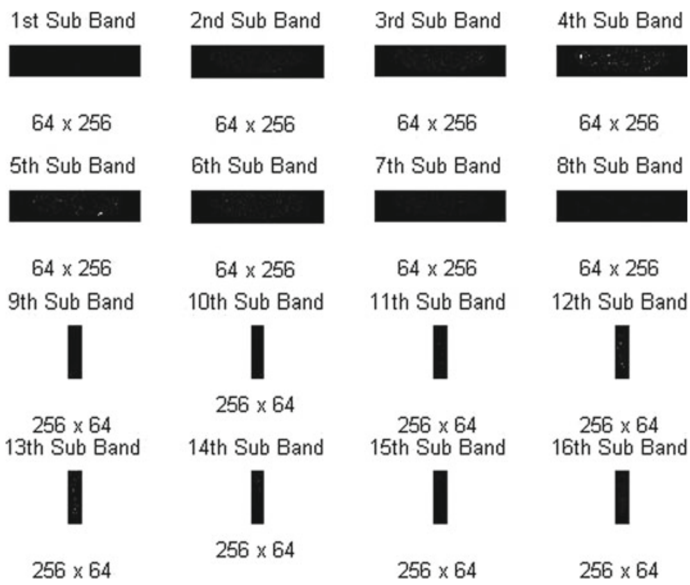


Fig. 3 A four-level Contourlet decomposition of input image from set 1

$$S_{\mathrm{BC}} = \sum\nolimits_{c} \sum\nolimits_{ci} (\overline{x_i} - \mu_c)(\overline{x_i} - \mu_c)^{\mathrm{T}} \tag{2}$$

where S_{WC} and S_{BC} represent the scatter matrices of within class and between class, respectively. C denotes the number of classes, μ_c is the mean of the class, and x_i is the mean of all classes with the objective function defined by

$$J = w^{\mathrm{T}} S_{\mathrm{BC}} w / w^{\mathrm{T}} S_{\mathrm{WC}} w \tag{3}$$

C. CLASSIFIER DESIGN: The proposed work utilizes a radial basis feed forward neural network (RBFN) with an input layer, a hidden layer, and an output layer. The input layer of network is set of l units, which accept l-dimensional feature vector. A model of the RBFN is shown in Fig. 4.

In this paper, the output layer has single output unit to determine whether the claimed user face is authentic or not. Connections between the input and hidden layers have unit weights and need not be trained. The output of an RBFN is given as

$$y(o) = g\left(\sum_{k=1}^{m} w_i h_i(o) + w_o\right) \tag{4}$$

where w_i is the weight and h_i is the activation function taken to be Gaussian in this paper. The number of nodes in the output layer is dependent on the number of classes taken.

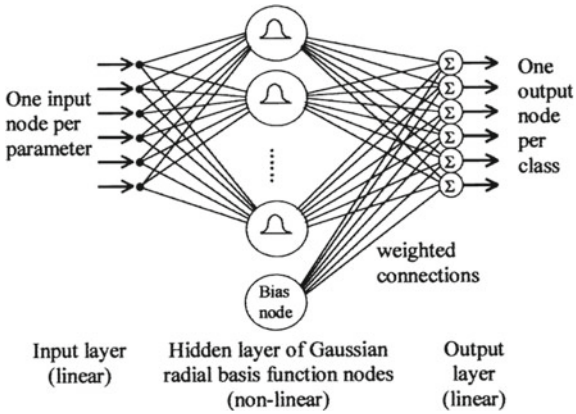


Fig. 4 Radial basis feed forward neural network model

4 Results and Discussion

In this paper, the input images have been taken from the ORL database and features extracted through the Contourlet transform and dimensions reduced using LDA. The feature set is fed into the RBF neural network and trained to obtain the classification as authentic or unauthentic. The data-base consists of 10 images of 40 different individuals taken at different angles, lighting conditions, and facial expressions. The input images are in .pgm format with sizes of 91 × 112 pixels taken in frontal and upright positions. The images taken for experimentation are depicted in Fig. 5.

The feature set obtained from the decomposition and LDA is fed into the neural network to obtain a high classification rate of nearly 98.8%. The proposed work has been compared with techniques like eigen-face, PCA, LDA-based face recognition, and wavelet-based recognition techniques, and results are tabulated. The efficiency of the proposed system is measured in terms of FAR (False acceptance Ratio] and false reject rate (FRR). False acceptance rate and false rejection rate are depicted as

$$\text{FAR} = N_\text{F}/N; \quad \text{FRR} = N_\text{R}/N \tag{5}$$

where N_F denotes the number of samples that were falsely accepted, while N denotes the total number of samples which were matched and N_R denotes the number of samples which were genuinely rejected. 200 images have been used for input out which 120 have been used for training the RBFNN. During feed forward, each input node receives an input value and relays it to each hidden neuron, which in turn computes the activation and passes it on to each output unit, which again computes the activation to obtain the net output. During training, the net output is compared with the target value and the appropriate error is calculated. From this, the error factor is obtained which is used to distribute the error back to the hidden layer. The weights are updated accordingly. In a similar manner, the error factor is calculated for units. The output layer contains one neuron. The result obtained from the output

Fig. 5 Input images taken from ORL database

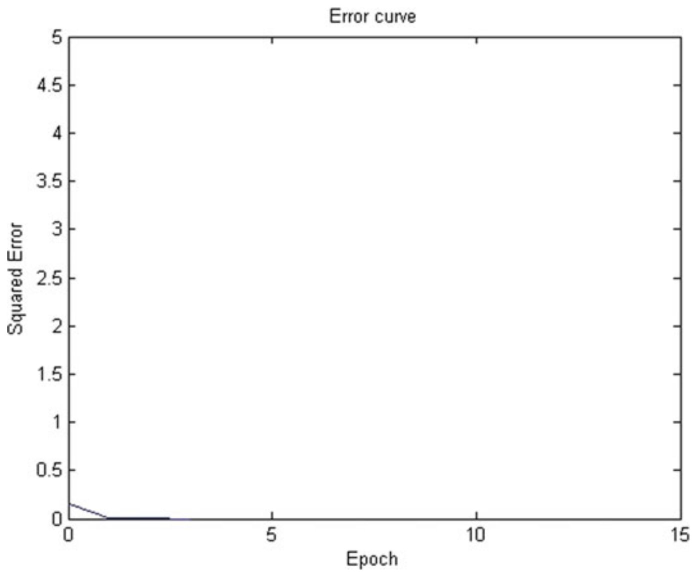


Fig. 6 Convergence of error versus number of epochs

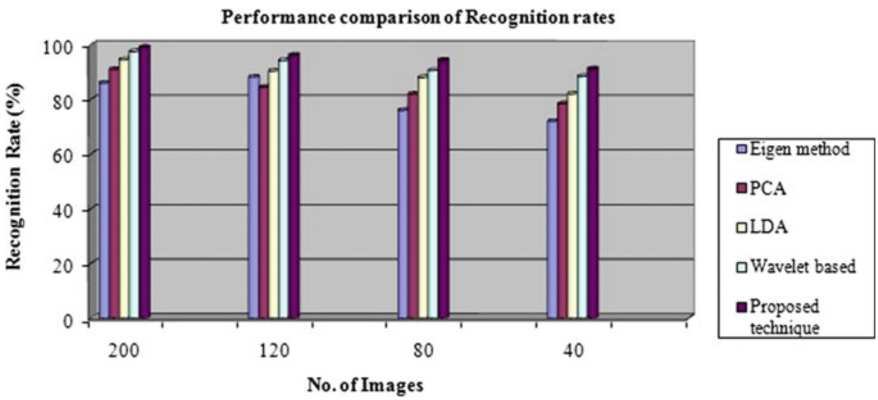


Fig. 7 Performance comparison of recognition rates

layer is given as the input to the RBF. The error incurred during the training phase in terms of epoch is shown in Fig. 6.

The proposed technique has been tested with a set of 200 images and quantified in terms of false acceptance rate and recognition rate. Figure 7 illustrates the comparative analysis. It could be clearly seen that the proposed technique outperforms the other conventional techniques by its recognition rate of nearly 99.2%. Out of the 200 images tested, three were found to be falsely accepted.

Table 1 Comparative analysis of proposed technique

No. of images	Acceptance ratio (%)					Execution time (s)				
	Eigen	PCA	LDA	Wavelet	Proposed technique	Eigen method	PCA	LDA	Wavelet	Proposed technique
200	92.5	93.8	95.14	95.89	97.25	80	78	77	72	50
120	90.1	90.5	91.1	92.4	94.2	64	63	61	57	42
80	86.5	88.4	89.25	90.1	93.1	51	48	47	41	38
40	82.4	86.4	88.4	89.2	90.89	44	40	41	36	32

Another important parameter is the computation time and as already mentioned in the previous sections, computation of eigen-faces takes longer which could be reflected in the observations tabulated in Table 1. Since the proposed technique does not involve computation of eigen faces, a drastic reduction in computation time could be observed. LDA provides the dimensionality reduction on the directional sub-bands, hence reducing the computational steps thus accounting for the reduced computation time. For a sample of 200 images, it could be seen that the classification results could be obtained in a little over 50 s.

5 Conclusion

The proposed frequency- domain- based dominant feature extraction algorithm provides an excellent space–frequency localization, which is clearly reflected in the high within-class compactness and high between-class reparability of the extracted features provided by the linear discriminant analysis algorithm. The use of a multi-resolution transform avoids the need to go for an entire image at a single instant to obtain the features. On the other hand, it provides a set of highly informative and directive sub-band to work with. LDA provides a good feature compaction reducing the dimension of the feature vector to over 1×100. The observed recognition rates indicate that the proposed technique is not sensitive to varying facial features due to lighting, environmental factors, and hence plays a key role in discriminating different faces. Moreover, it utilizes a very low-dimensional feature space, which ensures lower computational burden. For the task of classification, a radial-based classifier has been employed and it is found to provide a satisfactory recognition performance, hence reducing the complexity of the recognition network. From our extensive simulations on different standard face databases, it has been found that the proposed method provides high recognition accuracy of images with variations in orientations, lighting. This technique could be extended in future to moving images or images in moving background.

References

1. Bowyer Kevin W, Biswas Soma, Aggarwal Gaurav, Flynn Patrick J (2013) Pose-robust recognition of low-resolution face images. IEEE Trans Pattern Anal Mach Intell 35:3037–3049
2. Amani N, Shahbahrami A, Nahvi M (2013) A new approach for face image enhancement and recognition. Int J Adv Sci Technol 52(1):1–10
3. Al Allaf ONA (2014) Review of face detection systems based artificial neural network algorithms. Int J Multimedia Appl 6(1):1–16
4. Zhang D, Ding D, Li J, Liu Q (2015) PCA based extracting feature using fast fourier transform for facial expression recognition. In: Transactions on engineering technologies, pp 413–424
5. Patel Riddhi, Yagnik Shruti B (2013) A literature survey on face recognition techniques. Int J Comput Trends Technol 5(4):189–194
6. Gao Q, Gao F, Zhang H, Hao X-J, Wang X (2013) Two dimensional maximum local variation based on image Euclidean distance for face recognition. IEEE Trans Image Process 22(10):3807–3817
7. Park Usang, Jain Anil K (2010) Face matching and retrieval using soft biometrics. IEEE Trans Inf Forensics Secur 5:406–415
8. Yang MH, Kriegman DJ, Ahuja N (2002) Detecting faces in images: a survey. IEEE Trans Pattern Anal Mach Intell 24(1):34–58
9. Elbakry A (2002) Further experience with the tubularized-incised urethral plate technique for hypospadias repair. BJU Int 89(3):291–294
10. Matsugu M, Katayama T, Hatanaka K (2004) Method of extracting image from input image using reference image. U.S. Patent 6,766,055, Issued July 20
11. Li Z, Park U, Jain A (2011) A discriminative model for age invariant face recognition. IEEE Trans Inf Forensics Secur 6(3):1028–1037
12. Zheng Y, Zhang C, Zhou Z (2012) A wavelet based recognition for multispectral face. In: Proceedings of SPIE, vol 8401
13. Sharma P, Arya KV, Yadav RN (2013) Efficient face recognition using wavelet based generalized neural network. Mach Learn Intel Image Process 93(6):1557–1565
14. Anuradha K, Tyagi MK (2012) A novel method of face recognition using LBP, LTP and Gabor features. Int J Sci Technol Res 1(5):31–35
15. Moon H-M, Pan SB (2013) The LDA-based face recognition at a distance using multiple distance image. In: Seventh international conference on innovative mobile and internet services in ubiquitous computing, pp 249–255
16. Kaur A, Singh S, Taqdir A (2013) Face recognition using PCA and LDA techniques. Int J Adv Res Computer Commun Eng 4(3):308–310
17. Sahoolizadeh H, Jannesari A, Dousti M (2018) Noise suppression in a commongate UWB LNA with an inductor resonating at the source node. AEU-Int J Elect Comm 96:144–153
18. Kaushal A, Raina JPS (2010) Face detection using neural network & Gabor wavelet transform. Int J Comput Sci Technol (IJCST) 1(1):58–63
19. Abadi Mohammad et al (2011) Face detection with the help of Gabor wavelets characteristics and neural network classifier. Am J Sci Res 36:67–76
20. Bouzalmat A et al (2011) Face detection and recognition using back propagation neural network and Fourier Gabor filters. Signal Image Process Int J (SIPIJ) 2(3):15
21. Zhao Lina, Wanbao Hu, Cui Lihong (2012) Face recognition feature comparison based SVD and FFT. J Signal Inf Process 3:259–262
22. Do MN, Vetterli M (2001) Pyramidal directional filter banks and curvelets. In Image Processing, 2001. In: Proceedings 2001 International Conference on, vol 3, pp. 158–161. IEEE

Breast Cancer Recognition by Support Vector Machine Combined with Daubechies Wavelet Transform and Principal Component Analysis

Fangyuan Liu and Mackenzie Brown

Abstract The method of identifying the abnormal mammary gland tumor images was presented in order to assist the medical staff to find the patients with breast diseases accurately and timely. Db2 wavelet transform and principal component analysis (select the optimal threshold) is used to extract the effective features, support vector machine (set appropriate penalty parameter) is used to classify health and diseased samples, and 10-fold cross-validation is used to verify the classification result. The experimental results show that the method is feasible, the average sensitivity is $83.10 \pm 1.91\%$, the average specificity is $82.60 \pm 4.50\%$, and the average accuracy is $82.85 \pm 2.21\%$.

1 Introduction

The treatment of cancer is one of the most important problems for researchers in the present age. At present, the incidence of breast cancer is getting higher and higher, and the age span is getting wider and wider, which causes unpredictable damage to human health. In the light of breast cancer, experts from various industries combined computer-aided diagnosis with X-ray radiography technology or B-ultrasonic. Researchers can detect breast tumors at early stage, identify malignant and benign tumors accurately, and adopt effective treatment methods in a timely manner. There-

The original version of this chapter was revised: The Corresponding Author function has been assigned to Mackenzie Brown and the wrong labeling of images in figure 1 has been corrected. The correction to this chapter is available at https://doi.org/10.1007/978-3-030-00665-5_178

F. Liu
School of Computer Science and Technology, Nanjing Normal University, Nanjing 210023, Jiangsu, China
e-mail: 2191513453@qq.com

M. Brown (✉)
School of Engineering, Edith Cowan University, Joondalup, WA 6027, Australia
e-mail: mackbrown@ieee.org

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_177

fore, it is necessary and urgent to explore a kind of recognition method with good stability and high accuracy.

In response to the detection problem of human breast cancer, researchers have come up with a series of solutions. Xu [1] presented a kernel orthogonal transformation method based on linear orthogonal transform. In the experiment, the vector values of the test samples were compared with the set threshold after the transformation of the nuclear method to determine the malignant and benign breast tumors. By dividing the training–testing sets of samples in different proportions to improve the performance, the final experimental results show that the optimal partition detection rate of 4:1 reaches 100%. Senapati et al. [2] introduced a stable and effective way to identify breast cancer, which combined local linear wavelet neural network (LLWNN) with recursive least squares (RLS). The researchers used RLS in order to select the optimal parameters, including scale, translation, and local linear model. The accuracy of LLWNN-RLS is 97.2%, which is superior to other diagnostic methods. Uzer et al. [3] created two mixed feature selection methods to improve the comprehensiveness of the samples. The workers combined the sequential forward selection (SFS) and the sequential back selection (SBS) with PCA, respectively. Then, the artificial neural network based on SCG algorithm is used to classify the input features, and SBS-PCA+NN achieved 98.57% which is better than SFS-PCA+NN. Azar and El-Said [4] studied six kinds of detection methods based on support vector machine (SVM) for recognizing breast tumors: standard SVM, proximal SVM, smooth SVM, Lagrangian SVM, linear programming SVM, and finite Newton method for Lagrangian SVM. After adjusting the model parameters, the above methods obtained good recognition rate, in which the linear programming SVM performance is the best under each index.

Wavelet transform can extract important texture features of cancer foci. Hence, we used the wavelet transform [5–14] to extract distinguishing features. The principal component analysis was further used to reduce the features. Finally, support vector machine [15–20] was used to help establish the classifier.

We make a detailed exposition of the content after the introduction of the first section in this paper. The second section introduces the dataset and method used in the experiment, the third section analyzes the experimental results, and the fourth section summarizes the practicability and the extension of the proposed method.

2 Dataset and Method

We download public open-access mini-MIAS database [21], which contains 322 mammogram images sizes of 1024×1024. We picked up 200 samples randomly from the mini-MIAS dataset—100 healthy and 100 abnormal breast images, respectively (Fig. 1).

At the first step, we use discrete wavelet transform (DWT) and principal component analysis (PCA) to extract the features of breast images and select the most effective features. DWT is the discretization of the wavelet function in scale and dis-

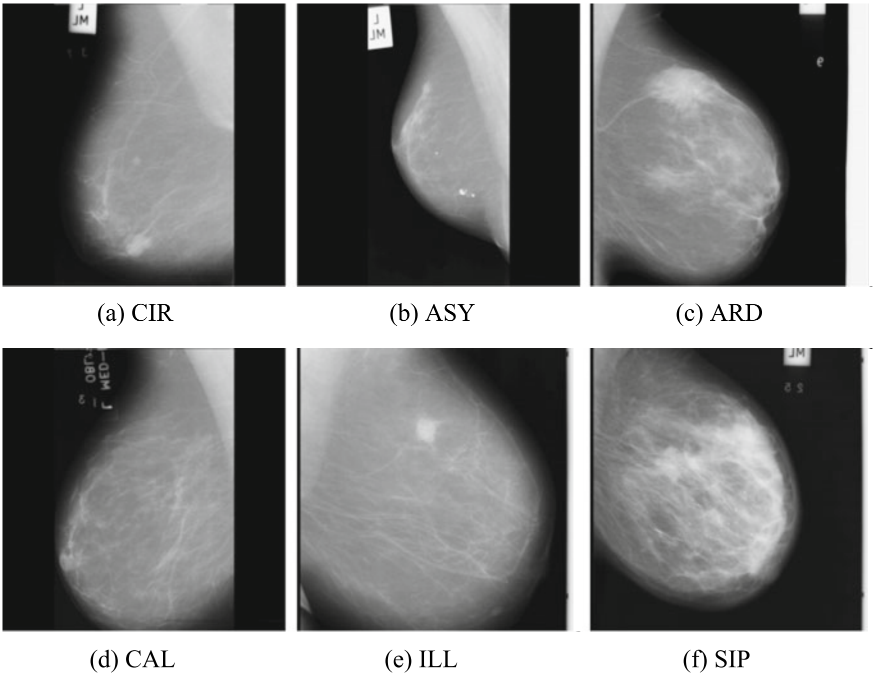


Fig. 1 Six abnormal breasts ((a) CIR = Circumscribed Mass; (b) ASY = Asymmetry; (c) ARD = Architectural distortion; (d) CAL = Calcification; (e) Ill = defined masses; (f) SPI = Spiculated masses)

placement, and the image of the mammary gland is decomposed, which is convenient for computer to process [22–24]. Furthermore, DWT can not only reduce the degree of redundancy of wavelet transform coefficients but also guarantee the integrality of information to some extent [25–29]. The expression of the wavelet basis function is shown in Formula (1), a represents the scale, τ represents the displacement, and t represents time. The Formulas (2) and (3) correspond to the discretization of scale and displacement, respectively. The expression of the discrete wavelet transform for function $f(t)$ is shown in Formula (4).

$$\psi_{a,\tau}(t) = \frac{1}{\sqrt{a}}\psi\left(\frac{t-\tau}{a}\right) \tag{1}$$

$$a = a_0^i,\ a_0 = 2,\ i = 0, 1, 2, \ldots \tag{2}$$

$$\psi_{a,\tau}(t) = \psi(t-\tau), a = 2^0, \tau = k\tau_0, k = 0, 1, 2 \ldots \tag{3}$$

$$WT_f(i,k) = \int_R f(t) \cdot \psi_{i,k}(t)d_t, \psi_{i,k}(t) = 2^{-\frac{i}{2}}\psi\left(2^{-i}t - k\tau_0\right) \tag{4}$$

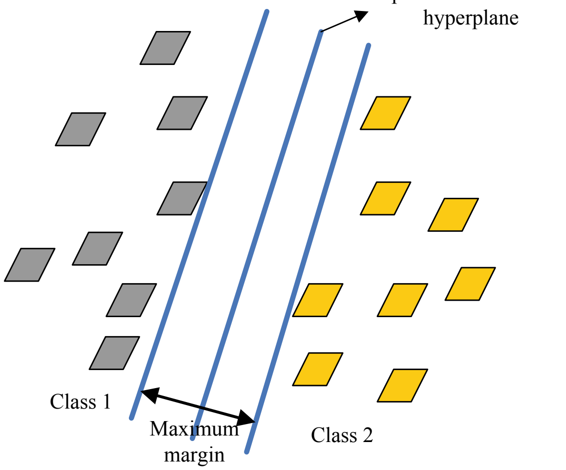


Fig. 2 Illustration of SVM

After DWT, the principal component analysis (PCA) was utilized to reduce the feature dimension and to remove the correlation between features [30]. The application of this method is considered because it can decrease computational complexity and improve the accuracy of identifying abnormal breast tumor images. The threshold *T* is an important parameter in PCA.

In the last step, we presented SVM to classify the breast tumor dataset to determine whether the breast is normal. SVM, as shown in Fig. 2, accurately divides the input eigenvectors according to the determination of the separating hyperplane and the maximum margin classification rule, which is mature in statistical classification and regression analysis. Therefore, SVM is used to identify medical image that can guarantee the reliability of the classification results. We did not use deep learning techniques, such as convolutional neural network [31–34] and autoencoder, because the dataset used in this study is relatively small.

Tenfold cross-validation was used to estimate this "DWT+PCA+SVM" approach. In tenfold cross-validation, nine folds were used as training set, and the rest was used as test set [35]. For a more strict analysis, this tenfold cross-validation repeats 10 times, and the average and standard deviation of the performances will be reported.

3 Experiments and Results

We used db2 wavelet in this study, the threshold of PCA is set to 95%, and the penalty parameter of SVM is set to 10. The classification performance measurement results of this 10×10-fold cross-validation are shown in Table 1. Here, we calculate the sensitivity (Sen), specificity (Spc), and accuracy (Acc) of each run at each fold.

Table 1 Measurement of proposed breast cancer detection at each fold (*Unit %*)

Sen	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10
R1	90	90	80	80	100	90	90	80	70	80
R2	80	70	90	70	90	80	80	70	90	90
R3	70	80	80	100	90	80	90	70	100	80
R4	80	100	80	100	90	70	70	90	90	80
R5	70	90	90	90	90	80	80	70	80	80
R6	90	70	100	90	60	90	80	80	80	90
R7	80	60	60	90	90	80	90	100	100	50
R8	70	90	80	70	90	80	80	90	90	80
R9	80	90	80	90	90	70	100	100	80	80
R10	50	90	90	90	70	80	90	90	100	80

(continued)

Table 1 (continued)

Spc	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10
R1	60	80	90	70	60	70	70	90	70	80
R2	70	90	80	80	90	80	80	100	100	80
R3	80	90	80	70	90	90	100	80	100	80
R4	70	90	90	90	100	90	90	80	70	70
R5	70	90	90	100	60	90	80	80	80	80
R6	60	80	80	70	80	70	90	100	80	70
R7	90	100	100	70	100	70	80	90	90	90
R8	60	70	80	100	80	80	90	70	80	70
R9	90	80	70	70	100	90	90	80	90	90
R10	90	90	90	50	100	80	90	80	100	90

(continued)

Table 1 (continued)

Acc	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10
R1	75	85	85	75	80	80	80	85	70	80
R2	75	80	85	75	90	80	80	85	95	85
R3	75	85	80	85	90	85	95	75	100	80
R4	75	95	85	95	95	80	80	85	80	75
R5	70	90	90	95	75	85	80	75	80	80
R6	75	75	90	80	70	80	85	90	80	80
R7	85	80	80	80	95	75	85	95	95	70
R8	65	80	80	85	85	80	85	80	85	75
R9	85	85	75	80	95	80	95	90	85	85
R10	70	90	90	70	85	80	90	85	100	85

Table 2 Average measure of proposed method (*Unit %*)

Run	Sen	Spc	Acc
1	85	74	79.50
2	81	85	83.00
3	84	86	85.00
4	85	84	84.50
5	82	82	82.00
6	83	78	80.50
7	80	88	84.00
8	82	78	80.00
9	86	85	85.50
10	83	86	84.50
Avr	83.10 ± 1.91	82.60 ± 4.50	82.85 ± 2.21

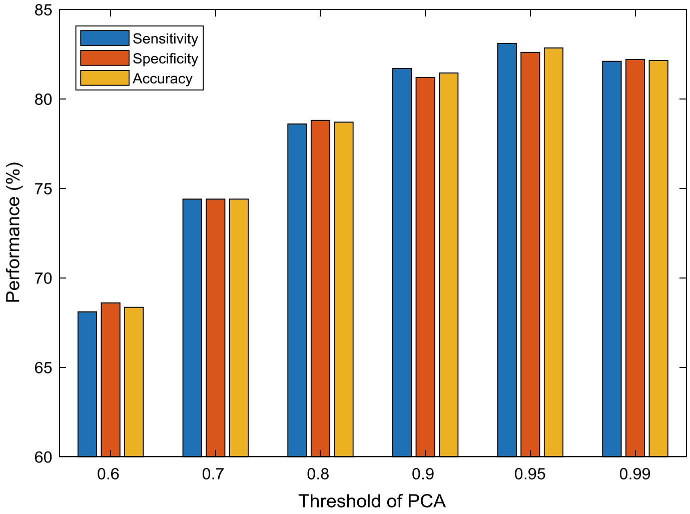


Fig. 3 Choosing optimal threshold of PCA

The average and standard deviation of sensitivity, specificity, and accuracy of our method are listed in Table 2, which indicates that the average sensitivity of our method is 83.10 ± 1.91%, the average specificity is 82.60 ± 4.50%, and the average accuracy is 82.85 ± 2.21%.

In the second experiment, we validated why we set the threshold of PCA to 95%. In Table 3, the effect of different PCA threshold values is shown. We compared the results at PCA threshold T of 60, 70, 80, 90, 95, and 99%, respectively. For a clear view, the comparison results are presented in Fig. 3. Obviously, $T = 95\%$ yields the greatest performance, and hence we chose this value as the optimal one in this study.

Table 3 Effect of different PCA threshold values (*Unit %*)

$T = 0.6$	Sen	Spc	Acc	$T = 0.7$	Sen	Spc	Acc	$T = 0.8$	Sen	Spc	Acc
R1	66	65	65.5	R1	71	75	73.0	R1	74	82	78.0
R2	76	73	74.5	R2	76	70	73.0	R2	84	76	80.0
R3	67	63	65.0	R3	77	74	75.5	R3	79	80	79.5
R4	67	68	67.5	R4	72	80	76.0	R4	82	72	77.0
R5	73	72	72.5	R5	78	78	78.0	R5	75	76	75.5
R6	64	66	65.0	R6	78	73	75.5	R6	83	81	82.0
R7	68	68	68.0	R7	75	75	75.0	R7	81	81	81.0
R8	67	68	67.5	R8	71	78	74.5	R8	75	84	79.5
R9	69	67	68.0	R9	70	70	70.0	R9	74	82	78.0
R10	64	76	70.0	R10	76	71	73.5	R10	79	74	76.5
Avr	68.10 ± 3.78	68.60 ± 3.95	68.35 ± 3.16	Avr	74.40 ± 3.10	74.40 ± 3.50	74.40 ± 2.17	Avr	78.60 ± 3.86	78.80 ± 3.99	78.70 ± 2.06

(continued)

Table 3 (continued)

$T=0.9$	Sen	Spc	Acc	$T=0.95$	Sen	Spc	Acc	$T=0.99$	Sen	Spc	Acc
R1	84	83	83.5	R1	85	74	79.5	R1	84	81	82.5
R2	80	84	82.0	R2	81	85	83.0	R2	79	80	79.5
R3	88	81	84.5	R3	84	86	85.0	R3	81	77	79.0
R4	83	82	82.5	R4	85	84	84.5	R4	78	81	79.5
R5	75	82	78.5	R5	82	82	82.0	R5	84	83	83.5
R6	84	83	83.5	R6	83	78	80.5	R6	85	86	85.5
R7	77	83	80.0	R7	80	88	84.0	R7	80	82	81.0
R8	88	74	81.0	R8	82	78	80.0	R8	82	79	80.5
R9	77	76	76.5	R9	86	85	85.5	R9	88	89	88.5
R10	81	84	82.5	R10	83	86	84.5	R10	80	84	82.0
Avr	81.70 ± 4.52	81.20 ± 3.43	81.45 ± 2.49	Avr	83.10 ± 1.91	82.60 ± 4.50	82.85 ± 2.21	Avr	82.10 ± 3.11	82.20 ± 3.49	82.15 ± 3.01

4 Conclusion

This study proposes a hybrid breast cancer recognition method based on db2 wavelet transform, principal component analysis, and support vector machine. This proposed method obtained promising results, with sensitivity of $83.10 \pm 1.91\%$, specificity of $82.60 \pm 4.50\%$, and accuracy of $82.85 \pm 2.21\%$.

The future studies contain applying advanced variants of wavelet transform, e.g., the stationary wavelet transform or the wavelet package transform. In addition, we may try to collect private breast image to increase the size of dataset, and then test deep learning methods, e.g., the autoencoders [36–38].

References

1. Xu Y, Zhu Q, Wang J (2012) Breast cancer diagnosis based on a kernel orthogonal transform. Neural Comput Appl 21(8):1865–1870
2. Senapati MR et al (2013) Local linear wavelet neural network for breast cancer recognition. Neural Comput Appl 22(1):125–131
3. Uzer MS, Inan O, Yilmaz N (2013) A hybrid breast cancer detection system via neural network and feature selection based on SBS, SFS and PCA. Neural Comput Appl 23(3):719–728
4. Azar AT, El-Said SA (2014) Performance analysis of support vector machines classifiers in breast cancer mammography recognition. Neural Comput Appl 24(5):1163–1177
5. Zhou XX, Zhang GS (2016) Detection of abnormal MR brains based on wavelet entropy and feature selection. IEEJ Trans Electr Electron Eng 11(3):364–373
6. Yang JQ et al (2016) A novel compressed sensing method for magnetic resonance imaging: exponential wavelet iterative shrinkage-thresholding algorithm with random shift. Int J Biomed Imaging. Article ID. 9416435
7. Sun P (2016) Preliminary research on abnormal brain detection by wavelet-energy and quantum-behaved PSO. Technol Health Care 24(s2):S641–S649
8. Yang M (2016) Dual-tree complex wavelet transform and twin support vector machine for pathological brain detection. Appl Sci 6(6): Article ID. 169
9. Zhou X-X (2016) Comparison of machine learning methods for stationary wavelet entropy-based multiple sclerosis detection: decision tree, k-nearest neighbors, and support vector machine. Simulation 92(9):861–871
10. Lu HM (2016) Facial emotion recognition based on biorthogonal wavelet entropy, fuzzy support vector machine, and stratified cross validation. IEEE Access 4:8375–8385
11. Nayak DR (2017) Detection of unilateral hearing loss by stationary wavelet entropy. CNS & Neurol Disord-Drug Targets 16(2):15–24
12. Wang S-H (2016) Single slice based detection for Alzheimer's disease via wavelet entropy and multilayer perceptron trained by biogeography-based optimization. Multimed Tools Appl. https://doi.org/10.1007/s11042-016-4222-4
13. Li Y, Cattani C (2017) Detection of dendritic spines using wavelet packet entropy and fuzzy support vector machine. CNS & Neurol Disord-Drug Targets 16(2):116–121
14. Li P, Liu G (2017) Pathological brain detection via wavelet packet tsallis entropy and real-coded biogeography-based optimization. Fundam Inform 151(1–4):275–291
15. Chen M (2016) Morphological analysis of dendrites and spines by hybridization of ridge detection with twin support vector machine. PeerJ 4:e2207
16. Chen S, Yang J-F, Phillips P (2015) Magnetic resonance brain image classification based on weighted-type fractional Fourier transform and nonparallel support vector machine. Int J Imaging Syst Technol 25(4):317–327

17. Dong Z (2014) Classification of Alzheimer disease based on structural magnetic resonance imaging by kernel support vector machine decision tree. Prog Electromagn Res 144:171–184
18. Ji G (2013) An MR brain images classifier system via particle swarm optimization and kernel support vector machine. Sci World J (130134)
19. Wu L (2012) Classification of fruits using computer vision and a multiclass support vector machine. Sensors 12(9):12489–12505
20. Wu L (2012) An MR brain images classifier via principal component analysis and kernel support vector machine. Prog Electromagn Res 130:369–388
21. The mini-MIAS database of mammograms (2018) Available from http://peipa.essex.ac.uk/info/mias.html
22. Gorriz JM (2017) Multivariate approach for Alzheimer's disease detection using stationary wavelet entropy and predator-prey particle swarm optimization. J Alzheimer's Dis https://doi.org/10.3233/jad-170069
23. Phillips P (2018) Intelligent facial emotion recognition based on stationary wavelet entropy and Jaya algorithm. Neurocomputing 272:668–676
24. Han L (2018) Identification of Alcoholism based on wavelet Renyi entropy and three-segment encoded Jaya algorithm. Complexity 2018(3198184):13
25. Zhou X-X et al (2016) Combination of stationary wavelet transform and kernel support vector machines for pathological brain detection. Simulation 92(9):827–837
26. Atangana A (2018) Application of stationary wavelet entropy in pathological brain detection. Multimed Tools Appl 77(3):3701–3714
27. Chen Y, Chen X-Q (2016) Sensorineural hearing loss detection via discrete wavelet transform and principal component analysis combined with generalized eigenvalue proximal support vector machine and Tikhonov regularization. Multimed Tools Appl 77(3):3775–3793
28. Chen Y, Lu H (2018) Wavelet energy entropy and linear regression classifier for detecting abnormal breasts. Multimed Tools Appl 77(3):3813–3832
29. Zhan TM, Chen Y (2016) Multiple sclerosis detection based on biorthogonal wavelet transform, RBF kernel principal component analysis, and logistic regression. IEEE Access 4:7567–7576
30. Schimit PHT, Pereira FH (2018) Disease spreading in complex networks: a numerical study with principal component analysis. Expert Syst Appl 97:41–50
31. Zhao G (2017) Polarimetric synthetic aperture radar image segmentation by convolutional neural network using graphical processing units. J Real-Time Image Process https://doi.org/10.1007/s11554-017-0717-0
32. Muhammad K (2017) Image based fruit category classification by 13-layer deep convolutional neural network and data augmentation. Multimedia Tools Appl. https://doi.org/10.1007/s11042-017-5243-3
33. Tang C (2017) Twelve-layer deep convolutional neural network with stochastic pooling for tea category classification on GPU platform. Multimedia Tools Appl. https://doi.org/10.1007/s11042-018-5765-3
34. Lv Y-D (2018) Alcoholism detection by data augmentation and convolutional neural network with stochastic pooling. J Med Syst 42(1):2
35. Gupta A, Kumar D (2018) Beyond the limit of assignment of metabolites using minimal serum samples and H-1 NMR spectroscopy with cross-validation by mass spectrometry. J Pharm Biomed Anal 151:356–364
36. Hou X-X (2017) Seven-layer deep neural network based on sparse autoencoder for voxelwise detection of cerebral microbleed. Multimedia Tools Appl. https://doi.org/10.1007/s11042-017-4554-8
37. Jia W (2017) Five-category classification of pathological brain images based on deep stacked sparse autoencoder. Multimedia Tools Appl. https://doi.org/10.1007/s11042-017-5174-z
38. Jia W (2017) Three-category classification of magnetic resonance hearing loss images based on deep autoencoder. J Med Syst 41(10):165

Correction to: Breast Cancer Recognition by Support Vector Machine Combined with Daubechies Wavelet Transform and Principal Component Analysis

Fangyuan Liu and Mackenzie Brown

Correction to: Chapter "Breast Cancer Recognition by Support Vector Machine Combined with Daubechies Wavelet Transform and Principal Component Analysis" in: D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30, https://doi.org/10.1007/978-3-030-00665-5_177

The Corresponding Author function has been assigned to Mackenzie Brown and the wrong labeling of images in figure 1 has been corrected.

The correction chapter and the book has been updated with the changes.

The updated original version of this chapter can be found at
https://doi.org/10.1007/978-3-030-00665-5_177

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_178