Lecture Notes in Electrical Engineering 222

Mohan S
S. Suresh Kumar
Editors

Proceedings of the Fourth International Conference on Signal and Image Processing 2012 (ICSIP 2012)

Volume 2

Springer

Lecture Notes in Electrical Engineering

Volume 222

For further volumes:
http://www.springer.com/series/7818

Mohan S · S. Suresh Kumar
Editors

Proceedings of the Fourth International Conference on Signal and Image Processing 2012 (ICSIP 2012)

Volume 2

Springer

Editors
Mohan S
Computer Science and Engineering
Dr. N.G.P. Institute of Technology
Coimbatore, Tamil Nadu
India

S. Suresh Kumar
Electronics and Communication Engineering
Dr. N.G.P. Institute of Technology
Coimbatore, Tamil Nadu
India

ISSN 1876-1100 ISSN 1876-1119 (electronic)
ISBN 978-81-322-0999-7 ISBN 978-81-322-1000-9 (eBook)
DOI 10.1007/978-81-322-1000-9
Springer New Delhi Heidelberg New York Dordrecht London

Library of Congress Control Number: 2012953282

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

Signal and Image Processing (SIP) is about the mathematical and computational methods that form the basis of the specialist modules covering the theory and application of SIP algorithms for the analysis, interpretation, and processing of data in diverse fields such as computer vision, robotics, acoustics, medical diagnosis, remote sensing, and telecommunications, and so on.

The SIP explores research and engineering issues related to the modeling of signals, developing and evaluating algorithms for extracting the necessary information from the signal, and the implementation of these algorithms. Specific research areas include filter design, fast transforms, adaptive filters, spectrum estimation and modeling, sensor array processing, image processing, motion estimation from images, speech processing, geophysics, computer-aided tomography, image restoration, robotic vision, and pattern recognition.

The 4th International Conference on Signal and Image Processing 2012 (ICSIP 2012), organized by Dr. N.G.P Institute of Technology, Coimbatore, India during Dec 13–15, 2012 provides academia and industry to discuss and present the latest technological advances and research results in the fields of theoretical, experimental, and application of signal, image, and video processing. ICSIP 2012 also aims to bring together engineers and scientists in signal, image, and video processing from around the world for better cohesion and collaboration in the fields of mutual interest.

The ICSIP 2012 called for a wide range of recent research papers in the field of SIP with applications, theories, and algorithms from all over the world from budding researchers. The submission was flooded with more than 250 papers from 12 different countries across the world. A total of 113 papers were selected based on blind review by experts from the respective fields for publication in the proceedings in two volumes by Springer India in Lecture Notes in Electrical Engineering.

The editors to express their gratitude to the founder of the ICSIP series Dr. P. Nagabhushan, Professor, University of Mysore, India for guidance right from the beginning. Also, the editors are thankful to Dr. S. Murali, President, Maharaja Institutions, Mysore, India for his valuable suggestions and timely input.

The editors are extremely thankful to all the keynote speakers (Prof. Kay Chen TAN-NUS, Singapore, Prof. Subhas, Massey University, New Zealand and Prof. P. Nagabhushan, UoM, India), tutorial speakers (Dr. Mathew Cherian-Kovai Medical Center and Hospital, India, Dr. Jharna Mazumdar, Former ADE DRDO, Ministry of Defence, India, Dr. Kumar Rajamani, GE, India, and Dr. Venkatesh Babu, IISc, Bangalore, India), panel members, reviewers, advisory, and organizing committee members. The editors convey their heartiest gratitude to Springer India team for accepting to publish the research findings in their Lecture Notes in Electrical Engineering, without whom this proceedings would not have been in existence today.

Above all, the editors are much thankful to the management of Dr. NGPIT, Coimbatore, India for hosting the prestigious conference ICSIP 2012. It was definite that the conference and the institution would complement each other in bringing high quality researchers in one place. The editors thank the Chairman- Dr. Nalla G. Palaniswami, Secretary- Dr. Thavamani D. Palaniswami, CEO- Dr. O. T. Buvaneswaran, Principal and all faculty members of Dr. NGP Institute of Technology, Coimbatore, who have been given great support by extending their seamless contribution to bring this proceeding in a big way.

Coimbatore, India

Mohan S
S. Suresh Kumar

About the Editors

Dr. Mohan S
B.E. (CSE), M.Tech. (CCT), Ph.D. (CS), MIEEE. With over 14 years of experience in Teaching, Research, and Industry (Infosys), Currently serving as Professor and Head, Department of CSE, Dr. NGP Institute of Technology, Coimbatore, Tamil Nadu, India. Part of organizing committees in ICCR 05 and 08, Organizing Chair in ICLDCM 2010.
Areas of interest: Computer Vision, Image/Video Processing
E-mail: s.mohan77@gmail.com

Dr. S. Suresh Kumar
Accomplished, Professor, Ph.D., C.Eng., MIEEE., MIEE., MIE., MAES., MISCA., MACCS., MSSI., MCSI., MISTE., with over 18 years of experience in education, research and development in a range of areas in Audio Signal Processing, Embedded Systems, and Power Electronics & Drives. Currently Director—Research at Dr. NGPIT, Coimbatore, Tamil Nadu, India.
E-mail: sskpsg@gmail.com

About ICSIP 2012

About the Proceedings

The proceedings include cutting-edge research articles from the Fourth International Conference on Signal and Image Processing (ICSIP), which is organized by Dr. NGP Institute of Technology, Kalapatti, Coimbatore. The Conference provides academia and industry to discuss and present the latest technological advances and research results in the fields of theoretical, experimental, and application of signal, image, and video processing.

The book provides latest and most informative content from engineers and scientists in signal, image, and video processing from around the world, which will benefit the future research community to work in a more cohesive and collaborative way.

Commitee Members

Patrons

Dr. Nalla G. Palaniswami, Chairman and Managing Director, Kovai Medical Center and Hospital, Coimbatore, India
Dr. Thavamani D. Palaniswami, Secretary, Dr. NGP Institute of Technology, Coimbatore, India
Dr. O. T. Buvaneswaran, CEO, KMCRET, Coimbatore, India
Dr. S. Dhandapani, Advisor, Dr. NGPIT, Coimbatore, India
Dr. K. Porkumaran, Principal, Dr. NGPIT, Coimbatore, India

Organizing Chairs

Dr. Mohan S, Professor and Head—CSE, Dr. NGPIT, Coimbatore
Dr. S. Suresh Kumar, Director-Research, Professor and Head—ECE, Dr. NGPIT, Coimbatore

Organizing Co-Chair

Prof. V. Vani, Professor and Head—IT, Dr. NGPIT, Coimbatore

Advisory Committee

Dr. Kay Chen Tan, National University of Singapore, Singapore
Dr. Subhas Chandra Mukhopadhyay, Massey University (Turitea), New Zealand
Dr. P. Nagabhushan, University of Mysore, Mysore, India
Dr. Murali S., Maharaja Institute of Technology, Mysore, India
Dr. Lakshminarayanan Samavedham, National University of Singapore, Singapore
Dr. R. M. Suresh, RMD college of Engineering, Chennai, India
Dr. Jharna Majumdar, Former Sc. G, DRDO, Bangalore, India
Dr. Mathew Cherian, KMCH, Coimbatore, India
Dr. R. Venkatesh Babu, SERC, IISc, Bangalore, India
Dr. Kumar Rajamani, GE Global Research, Bangalore, India

Organizing Committee Members

Faculty members of CSE, ECE and IT department, Dr. NGP Institute of Technology, Coimbatore, India

Contents

LVQ-Neural Network Based Signature Recognition System Using Wavelet Features

S. A. Angadi and Smita Gour

Abstract Signature recognition is an important requirement of automatic document verification system. Many approaches for signature recognition are presented in literature. A novel approach for off-line signature recognition system is presented in this paper, which is based on powerful wavelet features (maximum horizontal and vertical projection positions). The proposed system functions in three stages. Pre-processing stage; which consists of three steps: gray scale conversion, binarisation and fitting boundary box in order to make signatures ready for feature extraction, Feature extraction stage; where totally 64 wavelet based projection position features are extracted which are used to distinguish the different signatures. Finally in Neural Network stage; an efficient Learning Vector Quantization Neural Network (LVQ-NN) is designed and trained with 64 extracted features. The trained Neural Network is further used for signature recognition after the process of feature extraction. The average recognition accuracy obtained using this model ranges from 94 to 74 % with the training set of 15–50 persons.

Keywords LVQ-neural network · Wavelet features · Signature verification

1 Introduction

Signature recognition is one of the biometric authentication techniques, where the owner of the signature image is identified. Signature recognition can be classified into two main types, depending on the method of data acquisition: on-line and off-line signature recognition. In on-line recognition system, signature is obtained

S. A. Angadi (✉) · S. Gour
Department of Computer Science and Engineering, Basaveshwar Engineering College, Bagalkot, Karnataka, India

Mohan S. and S. S. Kumar (eds.), *Proceedings of the Fourth International Conference on Signal and Image Processing 2012 (ICSIP 2012)*, Lecture Notes in Electrical Engineering 222,
DOI: 10.1007/978-81-322-1000-9_1,

using an electronic tablet and other devices. Here, we can easily extract information about the writing speed, pressure points, strokes, acceleration as well as the static characteristics of signature data, using which the signature can be recognized. On other hand in off-line signature recognition, signature is available on a document, which is scanned to get the digital image representation. Processing these off-line signatures is complex and challenging due to the absence of stable dynamic characteristics. Recognizing a signature is very difficult due to highly stylish and unconventional writing skills, nature and variety of writing pen and non-repetitive nature of variation of the signatures because of age, illness, and geographic location.

Today an increasing number of transactions, especially related to financial and business are being authorized via signatures. A number of biometric techniques have been proposed for this authenticity verification. Among the vision-based ones, we can mention face recognition, iris scanning and retina scanning. But the equipment needed to obtain these biometrics may be too complex and expensive for large-scale deployment. But signature biometric authentication is rather simple and inexpensive and highly effective. Hence there is a need to have robust methods of automatic signature recognition for authenticity verification.

An exhaustive literature survey about signature verification has been carried out. Many researchers have followed different directions to achieve recognition of signature such as "*The compact three stage method of the signature recognition*" presented in [1], where personal signature is processed in the three stages (Preprocessing, Feature extraction, Comparison). In this method, the Hough transform is introduced, center of signature gravity is determined, and the horizontal and vertical signature histograms are obtained. Then personal signature is compared with the pattern from the signatures database. "*Extraction of global features for offline signature recognition*" is described in [2]. This work presents experimental method for the extraction of global handwritten signature features. The algorithm uses view-based approach and searches for the extreme values with the threshold value being applied. "*An Evolving Signature Recognition System"* reported in [4] has proposed a signature recognition method based on the fuzzy logic and genetic algorithm (GA) methodologies. "*A color Code Algorithm for Signature Recognition*" described in [5], deals with recognizing signature by morphological approach. Here dilation method is used repetitively to obtain the "Check pattern". This check pattern is obtained from the standard signature will be used for recognition of the signature. "*Off-line signature verification and recognition by support vector machine*" [8], *where* SVM (Support Vector Machine) is a new learning method introduced with a set of examples from two classes, a SVM finds the hyperplane, which maximizes the distance from either class to the hyper plane and separates the largest possible number of points belonging to the same class on the same side, etc.

Like this many more works [3, 6, 7, 9–18] have existed to solve the problem of signature recognition. Generally in every recognition system, there are three steps to be followed—image acquisition and preprocessing—feature extraction—recognition. All the suitable features have to be extracted from the digitized signature pattern and given to the suitable recognizer. But the fundamental problem in all the existing approaches is the extraction of important features from the signature.

In this work mainly maximum vertical projection position and maximum horizontal projection position features are extracted from the different regions of the image wavelets. "LVQ (Learning Vector Quantization) Neural Network" is used as a recognizer. This project has achieved the satisfying results using 64 wavelet features and LVQ-NN. As the number of persons increases the accuracy of a system is decreased. 94 % accuracy is obtained for 15 persons and for 50 persons, it has decreased to 74 %. 12 samples for each person are used for training the LVQ-NN and 4 samples completely different from training samples are used for testing. This accuracy is expected to increase if number of training samples increases.

The remaining part of the paper is organized into 5 sections. Section 2 discusses proposed methodology for signature recognition system using block diagram, flow chart and explanation of each block. In Sect. 3, the technique to extract features and type of features employed are discussed. In Sect. 4, the details about a signature recognizer used in this work is provided. In Sect. 5, the analyses of the result are presented. Section 6 concludes the work and lists some future work.

2 Proposed Methodology

The objective of this work is to build a robust signature recognition system. The methodology employed for this system is depicted in Fig. 1.

It involves two phases namely training phase and testing phase. The steps to be followed in both phases are as follows.

2.1 Image Acquisition

The image is acquired, by taking signatures from 50 members. Each member is given a white sheet with 16 rectangular boxes of 3 × 5 cm to give their signatures. Later these sheets are scanned and segmented to get individual signature image and stored in database as color image.

2.2 Preprocessing

To normalize the scanned signature images, some preprocessing steps have to be applied. The purpose in this phase is to make signatures to be of standard size and ready for feature extraction. In this work 3 different preprocessing techniques are

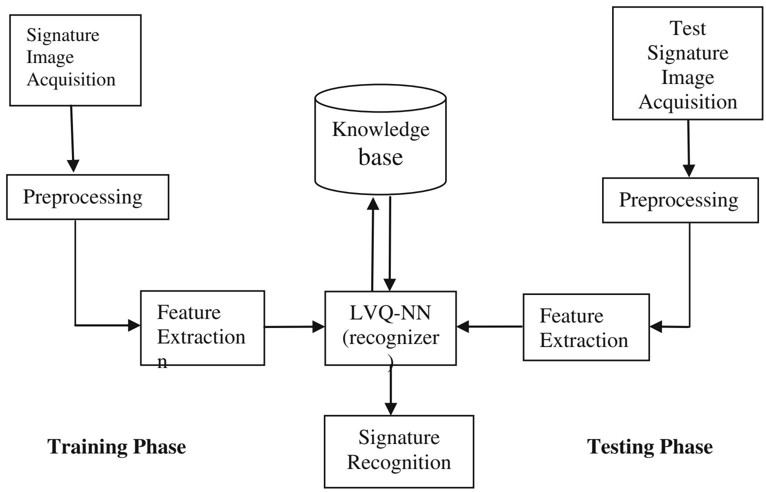


Fig. 1 Proposed Block Diagram for signatures recognition system

applied, namely -gray scale conversion -binarization -fitting boundary box which are discussed below.

2.3 Grayscale Conversion

Since the scanned images are stored in database as a color image, a three dimensional image (MXNX3) is not suitable for further processing, and should be converted into a grayscale image where each pixel is represented by a value in the range 0–255.

2.4 Binarization

It allows to reduce image information by removing background so that the image is black & white type. This type of image is much more easy for further processing.

2.5 Fitting Boundary Box

Individual signature images are automatically cropped to the size of the signature using fitting rectangular boundary box algorithm described in Algorithm 1, so that unnecessary areas are removed. It allows reducing the total number of pixels in the analyzed image (Fig. 2).

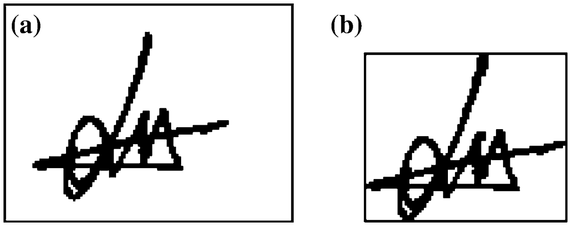


Fig. 2 **a** Image before fitting boundary box. **b** Image after fitting boundary

Algorithm 1: Fitting boundary box
Input: 8-bit binary image.
Output: Signature image reduced in size

- Start
- Step 1: Read signature image from specified location.
- Step 2: Find the size of an image (row, column).
- Step 3: Scan the image from top row, For (I = 1 to row) do the following
  - find the sum of *I*th row pixels
  - if sum is less than column then save that row number in k1 and stop scanning.
- Step 4: Scan the image from the bottom row, For (I = row to 1) do the following
  - find the sum of *I*th row pixels.
  - if sum is less than column then save that row number in k2 and stop scanning.
- Step 5: Scan the image from right most column, for (I = 1 to column) do the Following
  - find the sum of *I*th column pixels
  - if sum is less than row then save that column number in k3 and stop scanning.
- Step 6: Scan the image from left most column, for (I = column to 1) do the Following
  - find the sum of *I*th column pixels
  - if sum is less than row then save that column number in k4 and stop scanning.
- Step 7: Store the pixels of original image from row k1 to k2 and from column k3 to k4 in the variable I2.
- Step 8: Take I2 as a signature image reduced in size
- Stop

2.6 Feature Extraction

In this step, suitable wavelet features are extracted from the image. The procedure employed in this stage is described in the following. First- the image is divided vertically and horizontally into 8 equal sized blocks and diagonally into 4 blocks. Second- Discrete Wavelet Transform (DWT) is applied to each block to get 12*3 sub-images. Third- the features, maximum horizontal projection position and maximum vertical projection position are extracted from each of the sub images.

2.7 Training

This involves developing a suitable neural network model (LVQ-NN) and presenting extracted features to it iteratively so that neural network architecture is trained itself accordingly.

2.8 Testing

In testing, the features, maximum horizontal projection position and maximum vertical projection position are extracted from the testing image. These features are given to the trained LVQ-NN model, which classifies given sample and produces output to identify the owner of a signature.

The subsequent section of the paper describes the features employed, neural network used and the experimentation conducted.

3 Features Employed

To recognize the signature images, powerful features mainly maximum vertical projection position and maximum horizontal projection position are employed. The row, which has highest signature pixels, is taken as **maximum horizontal projection position**. This row number is normalized by last row number of sub image and taken as a feature of a signature. The column, which has highest signature pixels, is taken as **maximum vertical projection position**. This column number is normalized by last column number of sub image and taken as a feature of a signature.

These features are extracted using technique called wavelet. Mainly one level Discrete Wavelet Transform is applied on each of the signature image to decompose it into four sub-images. One of these sub-images is a smoothed version (approximation co-efficients) of the original image corresponding to the low pass

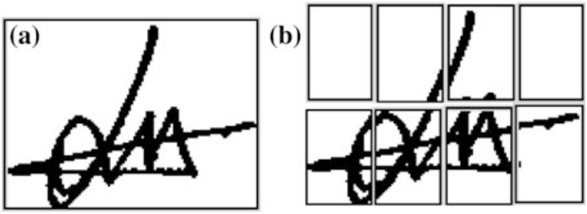


Fig. 3 **a** Preprocessed image. **b** 8 blocks of image

Fig. 4 **a** 3rd block of an image. **b** 4 sub images after applying DWT

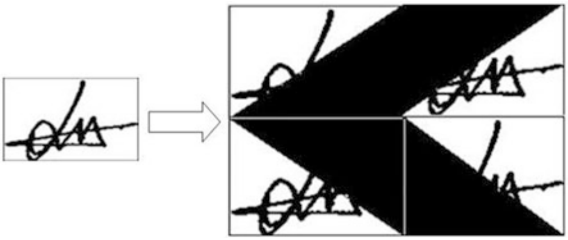


Fig. 5 Showing diagonally divided blocks of an image

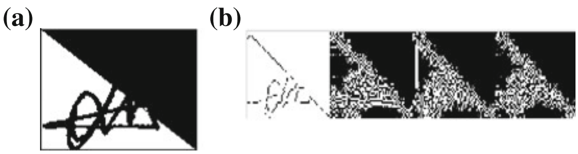


Fig. 6 **a** 3rd block of an image. **b** 4 sub images after applying DWT

information and the other three ones are high pass information (detailed co-efficients) that represent the horizontal, vertical and diagonal edges of the image respectively. The procedure followed in this work involves following steps.

First: The whole image is divided into 8 blocks as shown in the Fig. 3.

Table 1 MHP & MVP features extracted from 8 vertically divided blocks

Block 1	Vertical	MHP1 position, MVP2 position
	Horizontal	MHP3 position, MVP4 position
	Diagonal	MHP5 position, MVP6 position
Block 2	Vertical	MHP7 position, MVP8 position
	Horizontal	MHP9 position, MVP10 position
	Diagonal	MHP11 position, MVP12 position
Block 3	Vertical	MHP13 position, MVP14 position
	Horizontal	MHP15 position, MVP16 position
	Diagonal	MHP17 position, MVP18 position
Block 4	Vertical	MHP19 position, MVP20 position
	Horizontal	MHP21 position, MVP22 position
	Diagonal	MHP23 position, MVP24 position
Block 5	Vertical	MHP25 position, MVP26 position
	Horizontal	MHP27 position, MVP28 position
	Diagonal	MHP29 position, MVP30 position
Block 6	Vertical	MHP31 position, MVP32 position
	Horizontal	MHP33 position, MVP34 position
	Diagonal	MHP35 position, MVP36 position
Block 7	Vertical	MHP37 position, MVP38 position
	Horizontal	MHP39 position, MVP40 position
	Diagonal	MHP41 position, MVP42 position
Block 8	Vertical	MHP43 position, MVP44 position
	Horizontal	MHP45 position, MVP46 position
	Diagonal	MHP47 position, MVP48 position

Table 2 MHP & MVP features extracted from 4 diagonally divided block

Block 1	Vertical	MHP49 position, MVP50 position
	Horizontal	MHP51 position, MVP52 position
Block 2	Vertical	MHP53 position, MVP54 position
	Horizontal	MHP55 position, MVP56 position
Block 3	Vertical	MHP57 position, MVP58 position
	Horizontal	MHP59 position, MVP60 position
Block 4	Vertical	MHP61 position, MVP62 position
	Horizontal	MHP63 position, MVP64 position

Second: First level DWT with the filter called ‘db4′ is applied to each of the 8 blocks. DWT decomposes each block into four sub-images as shown in the Fig. 4.

Third: The original image is diagonally divided to get 4 blocks as shown in the Fig. 5.

Fourth: To each of the blocks, first level DWT is applied with the filter called ‘db4’ as done before, shown in the Fig. 6.

Fifth: For each of the detailed coefficient matrices, features such as maximum horizontal projection positions and maximum vertical projection positions are extracted.

All the 64 features extracted by the above procedure is listed in the Tables 1 and 2 MHP- Maximum Horizontal Projection MVP- Maximum Vertical Projection.

4 LVQ-Neural Network

The neural network used in this work is "Learning Vector Quantization (LVQ) neural network". LVQ can be understood as a special case of an artificial neural network, more precisely, it applies a winner-take-all Hebbian learning-based approach. It is a hybrid network uses both supervised and unsupervised learning. We have adapted LVQ-NN to recognize signatures because it uses unsupervised learning method, it takes less computation time and also is convenient to use. It has been used in many applications of pattern recognition and classification problems and has shown efficient results compared to other Neural Networks. To solve the problem of recognition/classification, it is to be trained which is discussed below.

4.1 Training of LVQ-NN

Presenting input vectors (features) iteratively and adjusting the location of hidden units based on their proximity to the input vector accomplishes training on LVQ network. Learning vector quantization employs a self-organizing network approach, which uses the training vectors to recursively "tune" placement of competitive hidden units that represent categories of the inputs. Once the network is trained, an

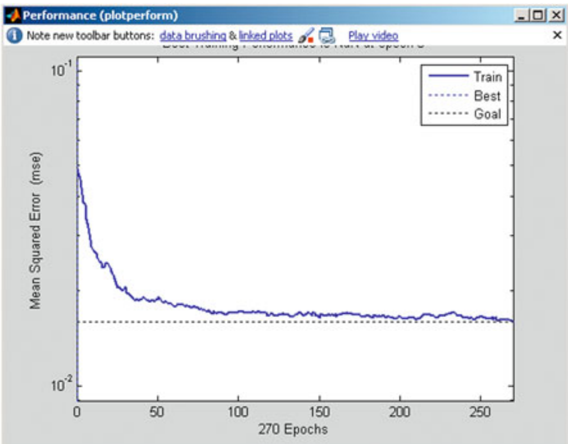
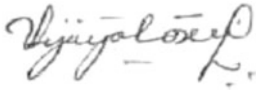


Fig. 7 Line graph showing performance goal met

Table 3 Testing results of 5 members among 15

Sl.No	Signature sample	Corresponding pattern	Owner of signature
1		100000000000000	S. A. Angadi
2		010000000000000	Vijayalaxmi
3		001000000000000	M. C. Elemmi
4		000100000000000	S. N. Benkikeri
5		000010000000000	S. S. Yendigeri

input vector is categorized as belonging to the class represented by the nearest hidden unit.

The performance goal set in this problem of signature recognition is 0.016 and number of hidden units taken is 60. The default learning rate that is 0.01 is used. The result of training is shown below in the Fig. 7.

4.2 *Testing LVQ-NN*

In testing, the features, maximum horizontal projection position and maximum vertical projection position are extracted from the test image. These features are given to the trained LVQ-NN model, which classifies given sample and produces output to identify the owner of a signature as shown in the Table 3.

5 Experimentation

Three separate experiments were conducted using the proposed model and results of the experiments are described in this section.

Experiment 1: In this experiment, the whole image is divided into 8 blocks and DWT is applied to each block to get 8*3 sub-images. Maximum horizontal projection position and maximum vertical projection position features are extracted from each of the sub images to get 8*3*2 features. This features vector was given to train the LVQ-NN. Accuracy obtained by this experiment was too low

Fig. 8 Accuracy graph

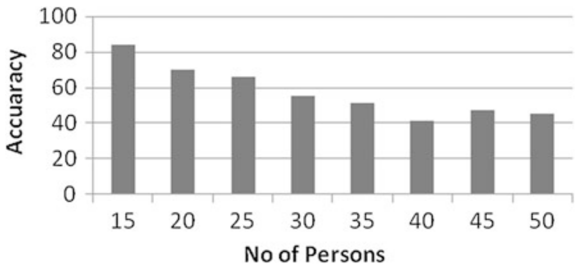


Table 4 Accuracy values

Persons	Accuracy (%)
15	84
20	70
25	66
30	55
35	51
40	41
45	47
50	45

Table 5 Accuracy values

Persons	Accuracy (%)
15	88
20	78
25	75
30	70
35	67
40	65
45	62
50	60

Fig. 9 Accuracy graph

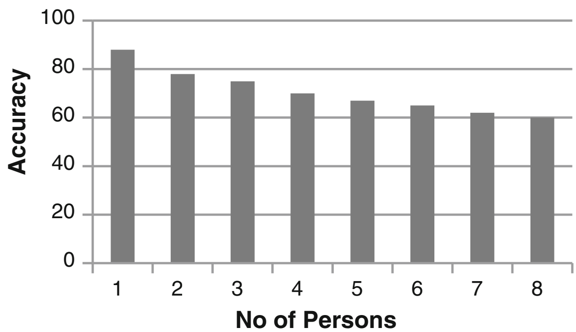


Table 6 Accuracy values

Persons	Accuracy (%)
15	94
20	91
25	88
30	85
35	80
40	77
45	76
50	74

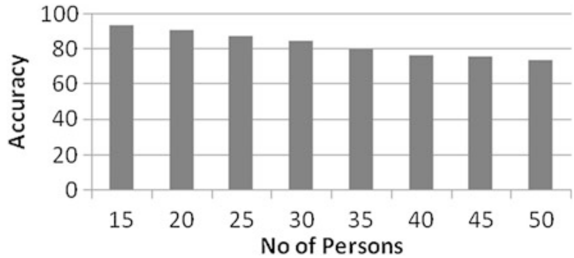


Fig. 10 Accuracy graph

(45–80 %) for 15–50 members, shown in the Fig. 8. 8 samples for each member were used for training (Table 4).

Experiment 2: In this experiment, another 16 features are added to the feature vector (48 + 16) obtained in the experiment 1. These 16 features are obtained by dividing the images diagonally into 4 blocks and following the same procedure as in experiment 1 (Table 5). Accuracy obtained by this experiment was improved (60–87 %) for 15–50 members shown in the Fig. 9.

Experiment 3: In this experiment, same feature vector obtained in the experiment 2 is used, but training samples of signature of each member is increased from 8 to 12 (Table 6). By this, accuracy is improved as expected (74–94 %) for 50–15 members as shown in the Fig. 10.

The experiments conducted indicate that as the number of different signatures increases the performance accuracy is decreased. Also with the increase of training samples the performance accuracy has been increased.

6 Conclusion

In this work, an off-line signature recognition system designed using 3 stages namely pre-processing, feature extraction and neural network stage in order to make the right decision is presented. This signature recognition is based on 64 powerful wavelet features of different signatures and the recognizer LVQ-NN. One of advantages of this LVQ-NN algorithm is short computational time. Experimental evidence has shown this method to provide substantial improvements. 74–94 % accuracy is obtained for the 50*12 signatures used for training LVQ-NN. This accuracy can be improved still with the more number of samples of each signature used for training. One problem faced, is the lack of samples needed to build a reliable signature recognition system and asses the performance. Unfortunately this system has failed to recognize the rotated signatures.

Future avenues of this work include an analysis of new features of signature image and combining those with the feature vectors used in this work to obtain better accuracy than the accuracy of present work.

References

1. Porwik Piotr, Będzińska (2007) The compact three stages method of the signature recognition. 6th Int Conf Computer Inf Syst Ind Manage Appl (CISIM'07) 0–7695–2894–5, 07 IEEE
2. Saeed K, Adamski M (2005) Extraction of global features for offline signature recognition
3. Kekre1 HB, Bharadi VA et al (2008) signature recognition by pixel variance analysis using multiple morphological dilations
4. Jayasekara B, Jayasiri A, Udawatta L (2006) An evolving signature recognition system. First Int Conf Ind Inf Syst ICIIS 8–11
5. Kulkarni VB (207) A colour code algorithm for signature recognition. Electron Lett Comput Vision Image Anal 6(1):1–12
6. Desira M (2008) Handwritten signature verification by independent component analysis
7. Igarza JJ, Hernáez I, Goirizelaia I, Espinosa K, Escolar J (2005) Off-line signature recognition based on dynamic methods. Proc SPIE 5779:336–343
8. Özgündüz E, Şentürk T, Karslıgil ME (2005) Off-line signature verification and recognition by support vector machine
9. Beheshti SGS (2009) Off-Line Persian Signature Identification and Verification Based on Image Registration and Fusion. J Multimedia 4(3):137–144
10. Chalechale A, Naghdy G, Premaratne P, Mertins A (2004) Cursive signature extraction and verification
11. Busch A, Boles WW (2002) Texture classification using multiple wavelet analysis. DICTA2002: Digital Image Comput Tech Appl 21–22
12. Rahul Rithe (2006) Fuzzy logic based off-line signature verification and forgery detection system 1–8
13. Sabourin R, Drouhard JP (1992) Offline signature verification using directional PDF and neural networks
14. Mohammed A, Abdala, Yousif NA (2009) Offline signature recognition and verification based on artifical neural network. Eng Tech J 27:7
15. Hasna JFA (2006) signature recognition using conjugate gradient neural networks. World academy of science engineering and technology 20
16. Pérez-Hernández A, Sánchez A (2004) Vélez JF Simplified stroke-based approach for off-line signature recognition
17. Endre Katona E (2012) Signature verification using neural nets
18. Jena D, Majhi B, Jena SK (2008) Improved offline signature verification scheme using feature point extraction method. J Comput Sci 4:2

Optimizing Dialogue Strategy in Large-Scale Spoken Dialogue System: A Learning Automaton Based Approach

G. Kumaravelan and R. Sivakumar

Abstract Application of statistical methodology to model dialogue strategy in spoken dialogue system is a growing research area. Reinforcement learning is a promising technique for creating a dialogue management component that accepts semantic of the current dialogue state and seeks to find the best action given those features. In practice, increase in the number of dialogue states, much use of memory and processing is needed and the use of exhaustive search techniques like dynamic programming leads to sub-optimal solution. Hence, this paper investigates an adaptive policy iterative method using learning automata that cover large state-action space by hierarchical organization of automaton to learn optimal dialogue strategy. The proposed approach has clear advantages over baseline reinforcement learning algorithms in terms of faster learning with good exploitation in its update and scalability to larger problems.

Keywords Human–computer interaction · Reinforcement learning · Learning automata · Spoken dialogue system

G. Kumaravelan (✉)
Department of Computer Science, Pondicherry Univeristy,
Karaikal Campus, Karaikal, India
e-mail: gkumaratcsbdu@gmail.com

R. Sivakumar
Department of Computer Science, AVVM Sri Puspam College,
Thanjavur, India
e-mail: rskumar.avvmspc@gmail.com

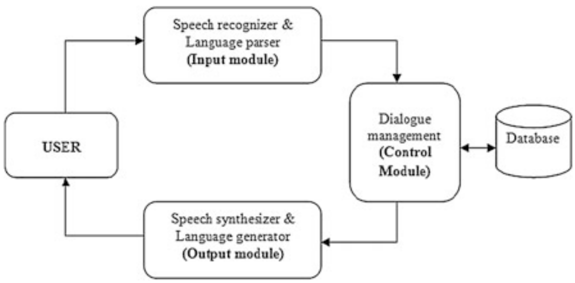
Mohan S. and S. S. Kumar (eds.), *Proceedings of the Fourth International Conference on Signal and Image Processing 2012 (ICSIP 2012)*, Lecture Notes in Electrical Engineering 222,
DOI: 10.1007/978-81-322-1000-9_2,

1 Introduction

Human–computer interfaces are now widely studied and have become one of the major interests among the scientific community. In particular, spoken dialogue system (SDS) is a natural language interface designed to make use of spoken language technology to accomplish a task between the user and a computer. Broadly, a SDS has three-modules, as shown in Fig. 1. The essential components are subsystems for input (conveying information from the user to the system), control (deciding how to react) and output (conveying information from the system back to the user) [1]. This paper is concern with the design of Dialogue Management (DM), the central component within the spoken dialogue systems to determine which communicative actions to take (i.e. what to say) given a goal and a particular set of observations about the dialogue history. In other words, they are responsible for controlling the flow of the interaction, sometimes referred to as dialogue strategy or policy in an efficient and natural way. This is a challenging task in most of the spoken dialogue systems wherein the dialogue strategy is handcrafted by a human designer which leads to errors, strenuous and non-portable.

Current research trends indicate attempts to find a way to automate the development of dialogue strategy using machine learning techniques. In practice, Reinforcement Learning (RL) techniques show appealing cognitive capabilities since they try to learn the appropriate set of actions to choose in order to maximize a scalar reward by following a trial and error interaction with an environment [2]. In this context, the dialogue strategy is regarded as a sequence of states with a reward for executing an action which in turn inducing a state transition in the conversational environment. The objective for each dialogue state is to choose such an action that leads to the highest expected long-term reward. For SDSs, these reward signals are associated with task completion and dialogue length. Hence, the system model covers the dynamics of Markov Decision Processes (MDPs) with a set of states S, a set of actions A, a state transition function, and a reward for each selected action. In this framework, a reinforcement learning agent aims at optimally mapping states to actions, i.e. "finding the optimal policy so as to maximize an overall reward" [3, 4].

Fig. 1 General architecture of a spoken dialogue system



However, the practical application of RL to optimize dialogue strategy faces a number of technical challenges such as choosing an appropriate reward function, scalability, robustness, and portability [5]. Several approaches to deal with the problem of large state-action spaces have been proposed in recent years. One of the approaches is based on the idea that not all state variables are relevant for learning a dialogue strategy and the state-action space is reduced by carefully selecting a subset of the available state variables by function approximation and hierarchical decomposition. If the relevant variables are chosen, useful dialogue strategies can be learnt. This technique has been applied successfully in several recent studies [6–9]. In addition, eXtended Classifier System (XCS) model has been applied in dialogue strategy optimization to evolve and evaluate a population of rules/and RL algorithm is applied to assign rewards to the rules [10]. However, it mitigates the curse of dimensionality problem by using a more compact representation with regions of state-action, but it finds less optimal solutions compared to tabular value functions.

The limitations of these contributions indicate that the exploration/exploitation trade-off in action selection strategy and curse of dimensionality in modeling the state space has to be solved completely. Most of the reinforcement learning based research attempt value iterative approach (Q-learning, SARSA) to find the optimal dialogue policy in action selection. However, when state-action spaces are small enough to represent in tabular form, Q-learning can be applied to generate a dialogue strategy. On the other hand, increasing the size of the state space for this algorithm has the danger of making the learning problem intractable referred to as "the curse of dimensionality." Another setback of the above baseline reinforcement learning algorithm is that it requires an update of the value function over the entire state space that is purely based on value iteration. In this case, one may get stuck on one iteration for a long time before any improvements in performance are made. Hence, tabular RL algorithms are designed to operate on individual state-action pairs with some practical limit to the size of the state-action table that can be implemented. Even with a relatively small number of state features and system actions, the size of the state-action space can grow very quickly. This constraint poses to be a problem for dialogue strategy developers. Hence, this paper proposes a scalable optimization approach which utilizes the policy iterative hierarchical structure learning automata algorithm of [10] to perform policy optimization over large state-action spaces.

2 Background

Reinforcement learning is a sub-area of Artificial Intelligence (AI) which considers how an autonomous agent acts through trial-and-error interaction with a dynamic unknown environment. Here, the agent refers to an entity that can perceive the state of the environment, and take actions to affect the environment's state. In turn, it receives a numerical signal called reinforcement from the environment for every action it takes. Its goal is to maximize the total reinforcements it receives over time. In reinforcement learning, an environment is often modeled as MDP, where the history of the environment can be summarized in a sufficient statistic called state to solve sequential decision making problems.

2.1 Dialogue as a MDP

One of the key advantages of statistical optimization methods for dialogue strategy design is that the problem can be formulated as a precise mathematical model which can be trained on real data. The Markov Decision Process (MDP) model serves as a formal representation of human–machine dialogue and provides the basis for formulating strategy learning problems. Every MDP is formally described by a finite state space *S*, a finite action set *A*, a set of transition probabilities *T* and a reward function *R*. At each time step t the dialogue manager is in a particular state. It executes the discrete action $a_t \in A$, transitions into the next state s_{t+1} according to the transition probability $p(s_{t+1}|s_t,a_t)$ and receives a reward r_{t+1}. In this framework, a DM is a system aiming at optimally mapping states to actions, that is finding best strategy π^* so as to maximize an overall reward *R* over time, i.e. the policy that selects those actions that yield the highest reward over the course of the dialogue.

2.2 Learning Automata

Learning Automata (LA) are adaptive decision-making devices operating on unknown random environment, and are associated with a finite set of actions and each action has a certain probability (unknown to the automaton) of getting rewarded by the environment of the automaton [11]. The aim is to learn the ways to choose the optimal action (i.e. the action with the highest probability of being rewarded) through repeated interaction on the system as shown in Fig. 2.

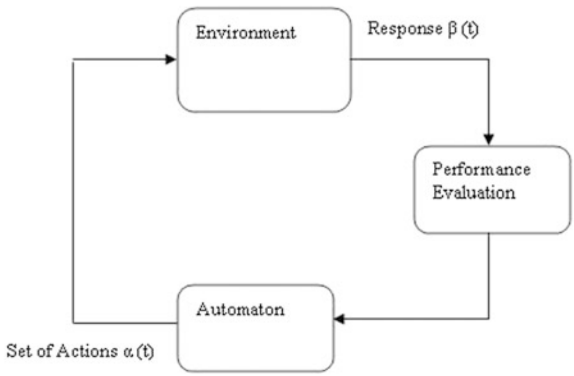


Fig. 2 Learning Automaton and its interaction with the environment

Formally, LA are represented by a triple $< \alpha, \beta, T >$, where α is the action set, β is the environment set and T is the learning algorithm. The learning algorithm is used to modify the action probability vector. The idea behind this update scheme T is that, when an action was successful, the action probability for the chosen action should be increased and all other action probabilities should be decreased appropriately. The general form is given by a following recurrence equations for the case where action $a(i)$ is selected at time step t (thus $a_t = a(i)$), the total number of actions is n and the reward obtained from the environment is r_{t+1}. The action probabilities are updated by the scheme given below:

$$p_{t+1}(i) = p_t(i) + \alpha r_{t+1}(1 - p_t(i)) - \beta(1 - r_{t+1})p_t \tag{1}$$

$$p_{t+1}(j) = p_t(i) - \alpha r_{t+1}p_t(j) + \beta(1 - r_{t+1})\left(\frac{1}{n - 1} - p_t(j)\right) \forall j \neq i \tag{2}$$

3 Methodology

A given dialogue task is decomposed into a root subtask and set of dialogue goals. Then, each dialogue goal is decomposed according to the nature of the slot filling strategy. Therefore, each dialogue sub-task in the hierarchy is represented with an MDP, and the hierarchy is denoted by $M = M_j^i$. The proposed method follows frame-based approach in modeling the dialogue structure in the form of frames that have to be filled by the user. Each frame contains slots that guide the user through the dialogue. In this approach, the user is free to take the initiative in the dialogue. The large body of transcribed and annotated conversation forms the basis for task identification, DA recognition, and form filling for task completion. In particular, this representation scheme classifies system actions in terms of their conversational domain, speech act, and task. For example, one possible system action is <about task, request info, dest city>, which corresponds to a system utterance such as What is your destination city?. The high-level structure of the dialogue manager is illustrated in Fig. 3, which includes several interconnected LA organised in hierarchical modules to handle such dialogue sub-tasks.

The dialogue policy is defined by a hierarchy of dialogue sub-tasks $\mathrm{M} = \mathrm{M}_\mathrm{j}^\mathrm{i}$ and that each sub-task can apply state abstraction to compress the state space. The indexes i and j only identify a subtask in a unique way in the hierarchy and they do not specify the execution sequence of subtasks because it is learnt by the learning automaton. Algorithm 1, elucidate the procedural form of Hierarchical Learning Automata (HLA) approach for handling knowledge-rich and knowledge-compact state space. HLA approach in practice receives dialogue subtask $\mathrm{M}_\mathrm{j}^\mathrm{i}$ and knowledge base k (based on domain artifacts) used to initialize the automaton at each level in the hierarchy and performs primitive action selections. But for composite actions it invokes recursively with a child subtask. When the subtask is completed

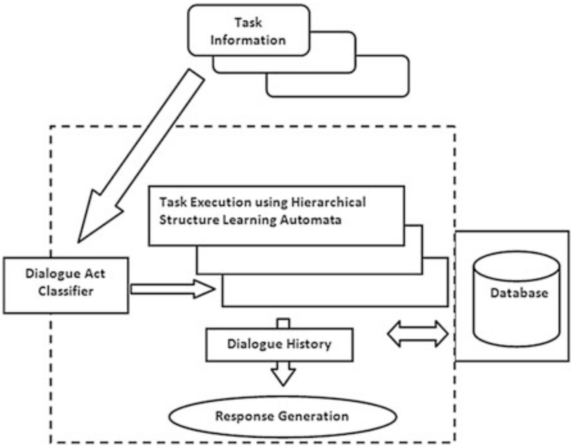


Fig. 3 Task hierarchy mechanism of the proposed method

with α time step it returns an average reward $R\left(a_t^l(h)\right)$ and continues its execution until finding a leaf state for the root subtask M_0^0. The algorithm is iterated until convergence occurs in optimal context-independent policies.

Algorithm 1 HLA-learning algorithm with knowledge compact states

```
Procedure HLA (Knowledge base k, Sub task M_j^i) return
averageReward R⟨a_t^l(h)⟩
Initialisation
  For all the learning automata: Initialize action
  probabilities:
                    ∀ a ∈ A : p(a) = 1/|A|;
  The estimates of all the actions ∀ a ∈ A : R⟨a_t^l(h)⟩ = 0;
  The estimates for the learning automata at level n
                    ∀ a ∈ A : L⟨a_t^l(h)⟩ = 0;
for each trial do
 Activate the top LA of the hierarchies
  for each level "l" in the hierarchy "h" do
   The active LA selects action a_t^l(h) probabilistically
   Perform joint-action selection    a = [a_t^l(1), ..., a_t^l(h)]
   Observe immediate reward r_t
   Compute R_t = r_t + γ L⟨a_t^l(h)⟩)
  end for
  for each level "l" in hierarchy "h "do
    Compute combined reward R⟨a_t^l(h)⟩ = R⟨a_t^l(h)⟩ + ρ[R_t −
                R⟨a_t^l(h)⟩]
    Update action probability p using R⟨a_t^l(h)⟩ as the
    reward for the L_RI scheme
    Propagates  r_{t+1} to the parent
  end for
end for
end procedure HLA
```

4 Experiments and Results

In this section, a slot-filling dialogue system based on the travel domain to verify the effectiveness of a dialogue strategy in a simulated learning environment for the proposed model is presented. The experimental design is based on the open agent architecture (OAA).The state space illustration of the chosen application has six slots representing all the currently available information regarding internal and external processes controlled by the dialogue system i.e., the knowledge of the concerned domain. For information seeking tasks, a common approach is to specify state variables on the number of slots that need to be filled and grounded. For example, a particular slot has not yet been stated (unknown), stated but not grounded (known), or grounded (confirmed). In addition, A record of the number of times each slot was asked for or confirmed by the system was necessary to indicate that a dialogue was not progressing sufficiently, perhaps due to the persistent (simulated) misrecognition of a particular slot. With these factors, the list of state variables and their possible values are given in Table 1.

The preferred system and user DAs which comprise action space for modeling dialogue strategies are summarized in Table 2. The system dialogue acts allow the system to request the user for the slot values and to restart or end the dialogue. Finally, the system can then present the results of a user's database query. The user dialogue acts allow the user to provide slot information, allow the user to terminate the dialogue, ask for help, and start the dialogue from the beginning.

Hence with 2,916,000,000 unique states and 10 system actions, the selected size of the state-action space explored by the experimental setup is 2.9 × 109. For tractable hierarchical learning, the state-action representation is decomposed into 7 sub-tasks. Figure 4 illustrates the sub-task hierarchy of the aforementioned application domain and Table 3 describes the state variables actions per subtask. It has been clearly emphasized that the each sub-tasks undergoes a state abstraction procedure by ignoring irrelevant variables.

Table 1 State variables representation in travel planning SDS

State variable	Possible values	State space size
dep_city_confidence	unknown, known, confirmed	3
dest_city_confidence	unknown, known, confirmed	9
date_confidence	unknown, known, confirmed	18
time_confidence	unknown, known, confirmed	324
brand_confidence	unknown, known, confirmed	972
location_confidence	unknown, known, confirmed	2,916
dep_city_times_asked	1..10	29,160
dest_city_times_asked	1..10	291,600
date_times_asked	1..10	2,916,000
time_times_asked	1..10	29,160,000
brand_times_asked	1..10	291,600,000
location_times_asked	1..10	2,916,000,000

Table 2 Common system and user dialogue acts used in travel planning SDS

System acts	User acts
Greeting	Command(bye)
Goodbye	Command(request_help)
Restart	Command(restart)
Request_Info(dep_city)	Provide_ Info(dep_city)*
Request_Info(dest_city)	Provide_ Info(dest_city)*
Request_Info(date)	Provide_ Info(date)*
Request_Info(time)	Provide_ Info(time)*
Request_Info(brand)	Provide_ Info(brand)*
Request_Info(location)	Provide_ Info(location)*
Database Results	Answer(yes); Answer(no)

* Multiple slot values can be provided in a single utterance

Table 3 State variables and actions of the subtask hierarchy in the travel planning system

Subtask	State variables	Actions
M_0^0	GIF, SAL, F0,H0	M_1^1, M_1^2, greeting(),
M_1^1	MAN, OPT	M_2^1, M_2^2, greeting(), restart(), database_results()
M_1^2	MAN, OPT	M_2^3, M_2^4, greeting(), restart(), database_results()
M_2^1	C00,C01,C02,C03	request_Info(dep_city) + imp_confirmation + exp_confirmation, provide_ Info(dep_city); request_Info(dest_city) +imp_confirmation + exp_confirmation, provide_ Info(dest_city); request_Info(date) + imp_confirmation + exp_confirmation, provide_ Info(date); request_Info(time) + imp_confirmation + exp_confirmation, provide_ Info(time);
M_2^2	C04	command(request_help); answer(yes);answer(no);
M_2^3	C05,C06	request_Info(brand) + imp_confirmation + exp_confirmation, provide_ Info(brand); request_Info(location) + mp_confirmation + exp_confirmation, provide_ Info(location)
M_2^4	C07	command(request_help)); answer(yes);answer(no);

The values of state variables includes Goal In Focus (GIF) = {0 = flight details, 1 = hotel details}, Salutation (SAL) = {0 = null, 1 = greeting, 2 = goodbye}, {MAN, OPT, F0, H0} ← {0 = unfilled sub-task, 1 = filled sub-task, 2 = confirmed sub-task}, slot in focus (Cij) ← {0 = unknown, 1 = known, 2 = confirmed}.

To assert the impact of the proposed method, different sizes of datasets (dialogue corpora) which represent the problem space are required. However, it is not possible to collect large amounts of data. Therefore, a user simulation technique [12] has been used to generate different datasets. The task of the user simulator is to provide examples of how a real user would behave while interacting with the system. It provides user actions at a dialogue act level and uses two main

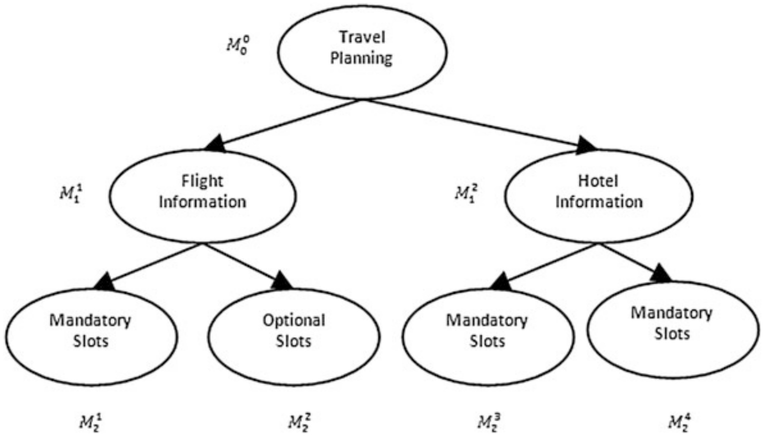


Fig. 4 The hierarchy of LA induced for the dialogue sub-tasks in the travel planning system

components user goal and user agenda. At the start of each dialogue, the goal describes the full set of constraints that the user requires to satisfy such as departure city, destination city, date and time. The agenda stores an ordered list of dialogue acts that the user is planning to use in stack like structure in order to complete its task. At the end of every dialogue episode the system receives a -1 penalty for every action it takes, a final reward of +20 in case of successful dialogue when all the necessary information has been obtained. In the case where no flight detail matches the attribute-slot values, the dialogue is deemed successful and a suitable alternative is offered. Since a characteristic dialogue will require about six or seven turns to complete, this implies that the achievable average reward has an upper bound of 14.

In each experiment, dialogue strategies were allowed to evolve over a fixed number of dialogues. The goal of the system is to acquire the values for the slots (attributes) with concern to each subtask (flight and hotel booking). In this case, the state-action space representation follows 2 hierarchies of 4 levels, with 10 actions per automaton. This gives a total of $(2^4)^{10} = 1.09 \times 10^{12}$ solution paths. Each action in the respective hierarchy is selected based on $L_{R\text{-}I}$ scheme as stated in Eqs. (1) and (2). The average reward is normalized to the interval [0–1] which determines the probability of action selection in each automaton in the hierarchy. The accuracy of the proposed approach is tested in function of the learning rate as shown in Fig. 5.

Figure 6 shows the reward value plotted against the number of iterations averaged over 10 training runs of 2000 episodes (or dialogues) by various approaches. With one time step of history, the baseline RL and HLA methods both appear to converge after about 500 dialogues. However the baseline RL method show clear signs of instability whereas HLA does not. This is due to the fact that the non-mobile LA defined in the levels of hierarchy does not move around the

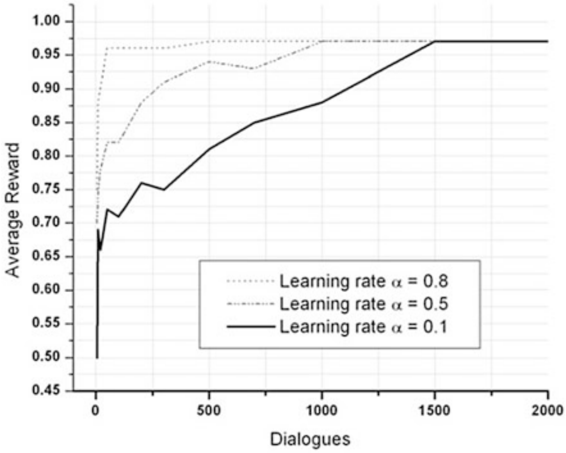


Fig. 5 The average reward using HLA method for various learning rates

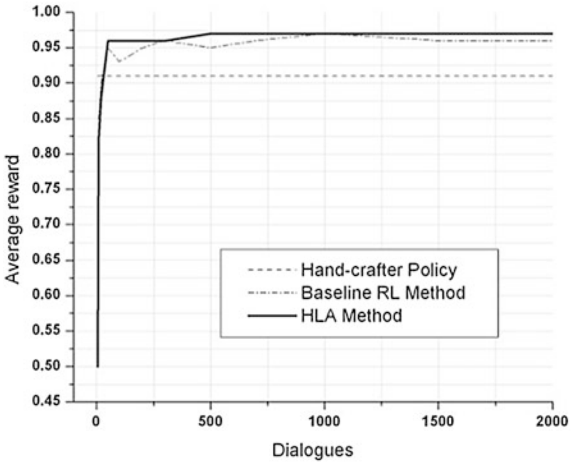


Fig. 6 Dialogue strategy learning by different approaches

state space but stay in their own state to get active and learn to take actions only in their own state of the MDP. However, in the case of baseline hierarchical learning approach using Q-learning algorithm, each Q-learner learns individually the associated Q-values with their own action rather than joint actions.

As a consequence the tabular approaches are completely independent and have no knowledge of the other Q-learner acting in the environment and influencing their reward. However, when the system has a unique limiting distribution over the state space, both independent LA and independent Q-learners are able to find a strategy for optimal action. Although the latter needs good exploration settings, it may take a very long time before convergence. In addition, the proposed approach appears quite stable and converges consistently with improved exploitation in its updates compared to the baseline tabular approach. This is one of the advantages of the HLA method when developing a new system or adapting it to changes in the tendency in the data.

5 Conclusion

This paper proposes learning dialogue strategies using policy iterative hierarchical structure learning automata that cover large state-action spaces under the formalism of Markov Decision Process in the simulated environment. In the analysis, it is found that the developed methodology is capable of learning optimal dialogue strategy in the context of large state-action spaces, application to travel planning domain. The hierarchical representation allowed the system to automatically generate specialised action that takes into account the current situation of the dialogue depending on the use of expected cumulative reward. Faster learning has become possible when the state space is being divided among multiple interconnected automata that can work independently to provide better convergence and knowledge transfer which provides the solution to learn about previous problems that can be re-used in new problems. For future work, we believe that our approach can also be extended in the direct context of Partially Observable MDPs which account for approximately linear running time of learning algorithms when the state space is continuous and directly incorporates uncertainty imposed to a noisy channel.

References

1. McTear M (2004) Spoken dialog technology: toward the conversational user interface. Springer, New York
2. Sutton RS, Barto AG (1998) Reinforcement learning an introduction. MIT Press, Cambridge
3. Levin E, Pieraccini R, Eckert R (2000) A stochastic model of human-machine interaction for learning dialog strategies. IEEE Trans Speech Audio Process 8(1):11–23
4. Singh S, Litman D, Walker M (2002) Optimizing dialogue management with reinforcement leaning: experiments with the NJFun system. J Artif Intell 16:105–133
5. Paek T, Pieraccini R (2008) Automating spoken dialogue management design using machine learning: an industry perspective. Speech Commun 50(8-9):716–729
6. Pietquin O, Dutoit T (2006) A probabilistic framework for dialog simulation and optimal strategy learning. IEEE Trans Audio Speech Lang Process 14(2):589–599
7. Henderson J, Lemon O, Georgila K (2008) Hybrid reinforcement/supervised learning of dialogue policies from fixed data sets. Comput Linguist 34(4):487–512
8. Cuayáhuitl H, Renals S, Lemon O, Shimodaira H (2010) Evaluation of a hierarchical reinforcement learning spoken dialogue system. Comput Speech Lang 24(2):395–429
9. Toney D, Moore J, Lemon O (2006) Evolving optimal inspectable strategies for spoken dialogue systems. In: Proceedings of HLT, pp 173–176
10. Baba N, Mogami Y (2006) A relative reward-strength algorithm for the hierarchical structure learning automata operating in the general nonstationary multiteacher environment. IEEE Trans Syst Man Cybern- Part B Cybern 36:781–794

11. Thathachar MAL, Sastry PS (2004) Networks of learning automata: techniques for online stochastic optimization. Kluwer, Norwell
12. Schatzmann J, Weilhammer K, Stuttle MM, Young S (2006) A survey of statistical user simulation techniques for reinforcement-learning of dialogue management strategies. The Knowl Eng Rev 21(02):97–126

Analysis of SVD Neural Networks for Classification of Epilepsy Risk Level from EEG Signals

R. Harikumar, C. Ganeshbabu, M. Balasubramani and P. Sinthiya

Abstract The Electroencephalogram (EEG) is a complex signal that indicates the electrical activity of brain. EEG is a signal that represents that effect of the superimposition of diverse processes in the brain. Epilepsy is a common brain disorder. Out of hundred one person is suffering from this problem. Here we study a novel scheme for detecting epileptic seizure and classifying the risk level from EEG data recorded from Epileptic patients. EEG is obtained by International 10–20 electrodes system. Singular Value Decomposition (SVD) is used for feature extraction. The efficacy of the above methods is compared based on the bench mark parameters such as Performance Index (PI), and Quality Value (QV). A group of twenty patients with known epilepsy findings are analyzed. It was identified that Elman neural network is a good post classifier in the optimization of epilepsy risk levels.

Keywords EEG signals · Singular value decomposition · Elman neural network · Epilepsy risk level · Seizure

R. Harikumar (✉) · C. Ganeshbabu
Department of ECE, Bannari Amman Institute of Technology, Sathyamangalam, Erode, India
e-mail: harikumarrajaguru@gmail.com

C. Ganeshbabu
e-mail: bits_babu@yahoo.co.in

M. Balasubramani · P. Sinthiya
Bannari Amman Institute of Technology, Sathyamangalam, Erode, India

Mohan S. and S. S. Kumar (eds.), *Proceedings of the Fourth International Conference on Signal and Image Processing 2012 (ICSIP 2012)*, Lecture Notes in Electrical Engineering 222,
DOI: 10.1007/978-81-322-1000-9_3,

1 Introduction

Epilepsy is a brain disorder in which clusters of nerve cells, or neurons, in the brain sometimes signal abnormally. Neurons normally generate electrochemical impulses that act on other neurons, glands, and muscles to produce human thoughts, feelings, and actions. In epilepsy, the normal pattern of neuronal activity becomes disturbed, causing strange sensations, emotions, and behaviour, muscle spasms, and loss of consciousness. During a seizure, neurons may fire as many as 500 times a second, much faster than normal. In some people, this happens only occasionally; for others, it may happen up to hundreds of times a day. About 25–30 percent of people with epilepsy still continue to experience seizures even with the best available treatment. One of the best methods of diagnosis of Epileptic seizure is through EEG signal monitoring. It is essential to classify the risk level of EEG signal where in the diagnosis is provided accordingly.

2 Methodology

The objective of this paper is to classify the epileptic risk level of EEG signal using SVD neural networks. Elman Neural Network is applied on the classified data to identify the optimized risk level which characterizes the patient's risk level. The EEG data used in the study were acquired from twenty epileptic patients who had been under the evaluation and treatment in the Neurology department of Sri Ramakrishna Hospital, Coimbatore, India. A paper record of 16 channel EEG data is acquired from a clinical EEG monitoring system through 10–20 international electrode placing method. With an EEG signal free of artifacts, a reasonably accurate detection of epilepsy is possible; however, difficulties arise with artifacts. This problem increases the number of false detection that commonly plagues all classification systems. With the help of neurologist, we had selected artifact free EEG records with distinct features. These records were scanned by Umax 6696 scanner with a resolution of 600 dpi.

2.1 Acquisition of EEG Data

Since the EEG records are over a continuous duration of about 30 s, they are divided into epochs of two second duration each by scanning into a bitmap image of size 400 $\times$ 100 pixels. A two second epoch is long enough to detect any significant changes in activity and presence of artefacts and also short enough to avoid any repetition or redundancy in the signal. The EEG signal has a maximum frequency of 50 Hz and so, each epoch is sampled at a frequency of 200 Hz. Each sample corresponds to the instantaneous amplitude values of the signal, totalling 400 values for an epoch.

2.2 Singular Value Decomposition

To extract the features of the seizure, we use a Singular Value Decomposition (SVD). The SVD method has been a valuable tool in signal processing and statistical data analysis. A SVD of an $M \times N$ matrix X, representing the TFD of the signal x, is given by

$$X = U\Sigma VT, \tag{1}$$

where $U(M \times M)$ and $V(N \times N)$ are orthonormal matrices, and Σ is an $M \times N$ diagonal matrix of singular values ($\sigma i \quad j = 0$ if $i \quad \_ = j$ and $\sigma 11 \geq \sigma 22 \geq \cdots \geq 0$). The columns of the orthonormal matrices U and V are called the left and right SVs, respectively. An important property of U and V is that they are mutually orthogonal. The singular values (σii) represent the importance of individual SVs in the composition of the matrix. SVD takes vectors in one space and transforms them into another space. Advantage of SVD is that it combines two different uncertainty representations into a metric as total uncertainty. SVD decomposes uncertainty measures (possibility, belief, probability etc.), combined as a collection of vectors of different units, into a principle space. We need this feature since our uncertainty measures cannot be added directly, they contain different units (epilepsy risk level codes). SVD has been applied successfully in many other technical disciplines as a tool to reduce coupled non linear behaviour to uncoupled collections of linear behaviour. The highest Eigen value obtained is considered as the pattern of the known patient's epilepsy risk level. Figure 1 shows the data segmentation from time series into data matrix.

The method employs data segmentation procedure where time series data is decomposed into blocks in order to extract principal components. The principal component features are concatenated, and then are fed into the SVD module.

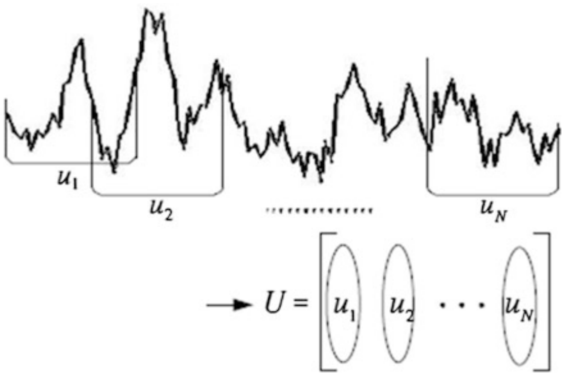


Fig. 1 Data segmentation: time series data is converted into data matrix

3 Elman Neural Networks as Post Classifier for Epilepsy Classification

Elman neural networks are a type of recurrent neural networks where connections between units form a directed cycle. A three layer network is used, with the addition of asset of context units. There are connections from the middle (hidden) layer to these context units with a weight of one. This feedback allows Elman networks to learn, recognize and generate temporal patterns, as well as spatial patterns. Every hidden layer is connected to only one neuron of the context layer through a constant weight of value one. Hence, the context layer constitutes a kind of copy the state of the hidden layer, one instant before. It is easy to find that the Elman network mainly consists of four layers: input layer, hidden layer, context layer and output layer as shown in Fig. 2. There are adjustable weights connecting each two neighbouring layers. Generally, it is considered as a special kind of feed forward neural network with additional memory neurons and local feedback. The self connections of the context nodes in the Elman network make it also sensitive to the history of input data which is very useful in dynamic signal modelling and analysis. Figure 3 shows the block diagram of Epilepsy risk level classification

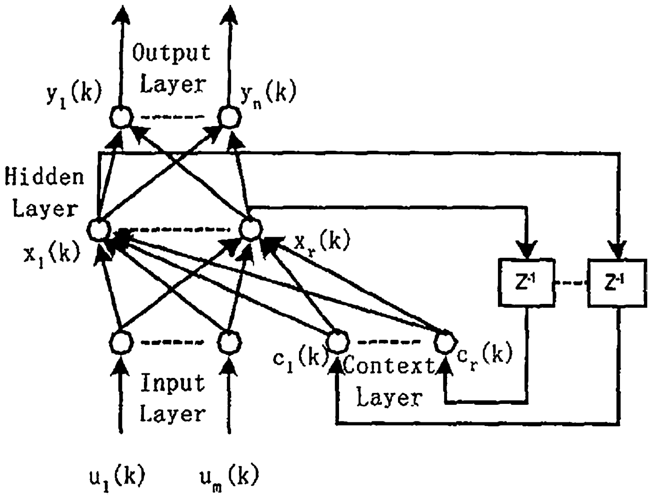


Fig. 2 Architecture of Elman neural network

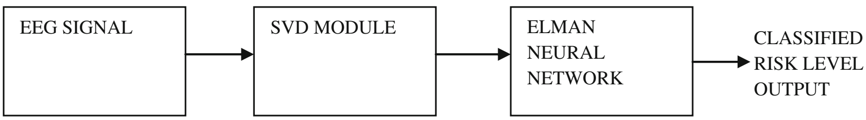


Fig. 3 SVD–neural network for epilepsy risk level classification

using Elman neural networks. The EEG signal is decomposed and normalised using SVD module and is optimised using the Elman neural network module.

3.1 Learning and Testing Procedures for the Selection of Elman Optimal Networks Architecture

The primary aim of developing an ANN is to generalize the features (epilepsy risk level) of the processed fuzzy outputs. We have used different architecture of Elman networks for optimization. The network is trained using LM (Levenberg-Maruardt) algorithm to minimize the square output error. This error back propagation algorithm is used to calculate the weights updates in each layer of the network. The simulations were realized by employing Neural Simulator 4.0 of Mat lab v.7.0 [18]. As the number of patterns in each database for training is limited, the technique of S-fold cross validation is employed to partition the data. The training process is controlled by monitoring the Mean Square Error (MSE) which is defined as

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(O_i - T_j)^2 \tag{2}$$

where O_i is the observed value at time i, T_j is the target value at model r; r = 1–10, and N is the total number of observations per epoch. As the number of hidden units is gradually increased from its initial value, the minimum MSE on the testing set begins to decrease. The optimal number of hidden units is that number for which the lowest MSE is achieved. A learning rate of 0.3 and a momentum term of 0.5 were used. Table 1 shows the estimation of MSE for the different Elman neural network architecture. Elman (20-20-20) is selected for its low Test MSE Index.

4 Results and Discussion

The SVD outputs in three epochs for each patient are optimized by the neural network approach as a single epileptic risk level. The relative performance of the neural networks is studied through the Performance Index and the Quality Value

Table 1 Estimation of MSE in various Elman network architectures

Architecture	Train MSE index	Test MSE index
20-20-20	3.3E-08	4.43E-08
16-16-16	4.21E-07	4.21E-07
4-4-4	3.4E-07	3.4E-07
8-8-8	0	0
16-2-16	0	2.94E-04

parameters. These parameters are calculated for each set of the patient and compared. Table 2 shows the normalised target output and the Mean square Error Function for Elman network Architecture (20-20-20).

The quality is determined by three factors.

(i) Classification rate
(ii) Classification delay
(iii) False Alarm rate

The quality value QV is defined as

$$Q_V = \frac{C}{\left(R_{fa} + 0.2\right) * \left(T_{dly} * P_{dct} + 6 * P_{msd}\right)} \tag{3}$$

where, C is the scaling constant, R_{fa} is the number of false alarm per set; T_{dly} is the average delay of the onset classification in seconds, P_{dct} is the percentage of perfect classification, and P_{msd} is the percentage of perfect risk level missed. A constant C is empirically set to 10 because this scale is the value of QV to an easy reading range. The higher value of QV, the better the classifier among the different classifier, the classifier with the highest QV should be the best. Table 3 shows the quality value obtained for average of twenty patients.

Table 2 Normalised target output and the mean square error function for Elman network architecture (20-20-20)

Normalised Target Output	MSE For Elman (20-20-20)
0.83915	2.116E07
0.82199	3.725E06
0.87011	3.349E06
0.90659	1.691E08
0.86502	2.161E06
0.99675	5.32E07
0.84192	1.69E08
0.92438	7.29E08
0.77743	3.028E06
0.86408	2.916E07
0.98164	2.496E06

Table 3 Quality value for average of twenty patients

Parameters	Elman neural network
Risk level classification rate (%)	98.92
Weighted delay (s)	1.978
False alarm rate/set	0.0108
Performance index (%)	98.92
Quality value	23.98

5 Conclusion

This paper analyses the performance of neural network in optimizing the epilepsy risk level of epileptic patients from EEG signals. Elman Neural network optimization technique is used to optimize the risk level. The classification rate of epilepsy risk level of above 93 % is possible in our method. The missed classification is almost nil for a short delay of 2 s. From this method we can infer the occurrence of High-risk level frequency and the possible medication to the patients. Also optimizing each region's data separately can solve the focal epilepsy problem. The future research is in the direction of a comparison between non heuristic optimization models with neural networks.

Acknowledgments The authors express their sincere thanks to the Management and the Principal of Bannari Amman Institute of Technology, Sathyamangalam for providing the necessary facilities for the completion of this paper. **This research is also funded by AICTE RPS.:F No 8023/BOR/RID/RPS-41/2009-10, dated 10th Dec 2010.**

References

1. Harikumar R, GaneshBabu C, Vijayakumar T (2012) Performance analysis of elman neural networks as post classifiers for wavelet transforms based feature extraction using hard and soft thresholding methods in the classification of epilepsy risk levels from EEG signals. Eur J Sci Res 71(2):221–232
2. Mizrahi EM, Kellaway P (1998) Neonatal electrocephalography, in diagnosis and management of neonatal seizures. Lippincott-Raven, Philadelphia, pp 99–143
3. Dingle AA et al (1993) A multistage system to detect epileptic form activity EEG. IEEE Biomed Eng 40(12):1260–1268
4. Mizrahi EM, Plouin P, Kellaway P (1997) Neonatal seizures, in Epilepsy. In: Engel J, Pedley TA (eds) A comprehensive textbook, vol 1, Chapter 57. Lippincott–Raven, Philadelphia, pp 647–663
5. Yuan Y (2010) Detection of epileptic seizure based on EEG signals. In: Proceedings of the IEEE EMBS sponsored 3rd international congress on image and signal processing (CISP 2010), pp 4209–4211
6. Sukanesh R, Harikumar R (2006) A simple recurrent supervised learning neural network for classification of epilepsy risk levels from EEG signals. IE India J Interdisc Panels 87(2):37–43
7. Harikumar R, Sukanesh R, Bharthi PA (2005) Genetic algorithm optimization of fuzzy outputs for classification of epilepsy risk levels from EEG signals. IE India J Interdisc Panels 86(1):9–17
8. Mirzaei A, Ayatollahi A, Gifani P, Salehi L (2010) EEG analysis based on wavelet- spectral entropy for epileptic seizures detection. In: Proceedings of the 3rd international conference on biomedical engineering and informatics (BMEI 2010), Changai, pp 878–882
9. Haselsteiner E, Pfurtscheller G (2000) Using time-dependent neural networks for EEG classification. IEEE Trans Rehabil Eng 8(4):457–463
10. Xanthanopoulus P et al (2010) A novel wavelet based algorithm for spike and wave detection in absence of epilepsy. In: Proceedings of the IEEE international conference on bioinformatics and bio engineering, pp. 14–19

11. Tzallas AT, Tsipouras MG, Fotiadis DI (2007) A time-frequency based method for the detection of epileptic seizure in EEG recording. In: Proceedings of the 12th IEEE international symposium on computer based medical systems (CBMS'07), pp 23–27
12. Kozek W, Hlawatsch F, Kirchauer H, Trautwein U (1994) Correlative time frequency analysis and classification of nonstationary random processes. In: Proceedings of the IEEE-SP international symposium on time-frequency and time-scale analysis, pp 417–420
13. Kandel ER, Schwartz JH, Jessell TM (1991) Principles of neural science, 3rd edn. Elsevier/North-Holland, New York
14. Nakamura A, Sugi T, Ikeda A, Kakigi R, Shibasaki H (1996) Clinical application of automatic integrative interpretation of awake background EEG: quantitative interpretation, report making, and detection of artifacts and reduced vigilance level. Electroencephalogr Clin Neurophysiol 98:103–112
15. Mormann F, Andrzejak RG, Elger CE, Lehnertz K (2007) Seizure prediction: the long and winding road. Brain 130:314–333
16. Van Drongelen W, Nayak S, Frim DM et al (2003) Seizure anticipation in pediatric epilepsy: use of Kolmogorov entropy. Pediatr Neurol 29:207–213
17. Lehnertz K, Mormann F, Kreuz T et al (2003) Seizure prediction by nonlinear EEG analysis. IEEE Eng Med Biol Mag 22:57–63
18. Firpi H, Smart O, Worrell G, Marsh E, Dlugos D, Litt B (2007) Highfrequency oscillations detected in epileptic networks using swarmed neural-network features. Ann Biomed Eng 35:1573–1584
20. Staba RJ, Wilson CL, Bragin A, Fried I, Engel JJ (2002) Quantitative analysis of high-frequency oscillations (80–500 Hz) recorded in human epileptic hippocampus and entorhinal cortex. J Neurophysiol 88:1743–1752
20. Rajna P, Clemens B, Csibri E et al (1997) Hungarian multicentre epidemiologic study of the warning and initial symptoms (prodrome, aura) of epileptic seizures. Seizure 6:361–368

Performance Analysis of an Activity Based Measurement of Blood Flow Using Impedance Plethysmography

R. Hari Kumar, C. Ganeshbabu, P. Sampath and M. Ramkumar

Abstract The project is mainly determined on the activity based measurement of the blood flow in humans. The objective of the project is to measure the blood flow in human limbs and to study the blood flow characteristics using an innovative methodology called Impedance plethysmography. It is also called as impedance test or blood flow or impedance phlebography. It can be used to measure arterial volume change that occurs with propagation of the blood pressure pulse in a limb segment. For this measurement, we assume a constant value of blood resistivity. However, blood resistivity may change under both physiological and pathological conditions. It is a non-invasive test that uses electrical monitoring in the form of resistance (impedance) changes to measure blood flow in veins of the leg. Information from this test helps doctors detect deep vein thrombosis. Using conductive jelly, the examiner strategically places four electrodes on the patient's calf. These electrodes are used to measure the impedance of the body and it is amplified. It is used to detect blood clots lodged in the deep veins of the leg, Screen patients who are likely to have blood clots in the leg, Detect the source of blood clots in the lungs (pulmonary emboli). It measures the blood volume changes in the human physiology which is used to detect the cardiovascular problems in the human.

Keywords Non-invasive · Cardiovascular · Plethysmography · Vein thrombosis · Thrombophlebitis

R. Hari Kumar (✉) · C. Ganeshbabu · P. Sampath · M. Ramkumar
Bannari Amman Institute of Technology, Sathyamangalam, Erode, India
e-mail: harikumarrajaguru@gmail.com

C. Ganeshbabu
e-mail: bits_babu@yahoo.co.in

M. Ramkumar
e-mail: ramcom09@gmail.com

Mohan S. and S. S. Kumar (eds.), *Proceedings of the Fourth International Conference on Signal and Image Processing 2012 (ICSIP 2012)*, Lecture Notes in Electrical Engineering 222,
DOI: 10.1007/978-81-322-1000-9_4,

1 Introduction

In earlier days, External transducers were used for various measurements but at times instead of an external transducer, the living system itself can be made to modify an externally applied electrical signal. For such measurement, the living system may be termed as an active transducer. A representative example of active transducer measurement is electrical impedance plethysmography. This technique has been applied to the measurement of the blood content of fingers and limbs, measurement of volume changes of the lungs during breathing and estimation of changes in the stroke volume. The basic principle underlying the Plethysmography measurements are that the impedance of the body segments under study changes during respiration and cardiac activity. Fundamental principles underpinning the study of cardiovascular physiology can be emphasized by measuring blood flow and also measure the resistance offered by the body to the flow of blood and in turn we can also estimate the diabetes level, which is the major contribution from our project. Measurement of blood flow provides valuable information about function and regulation of the circulation. Blood flow (Q) is dependent on and resistance (R) to the flow of blood. Therefore, measurements of blood flow can reflect changes in cardiac output, mediated by heart rate and/or stroke volume, although in some circumstances regional blood flow may change independently of any alteration in cardiac output. Such physiological changes are rapidly evoked in response to a variety of stresses, including physical exercise, drugs, changing posture, and local or systemic heating and cooling, that may be used to examine homeostatic control mechanisms. The Plethysmography is essentially a volume recorder that detects subtle changes in the volume of an organ over time. In this project we convert this into a function of resistance and deduce the other parameters that are essential for the measurement of blood flow. This Resistance undergoes minute changes when blood flow within the body changes. All these measurement are taken from a number of people and then we relate it to their corresponding body weights and physique, if the values are not found to be optimum then the patient is found to be having diabetes or not.

1.1 Motivation

This method is very much helpful for biomedical purposes, because Plethysmography techniques do not necessitate any surgical work. This technique is used for measurement of blood flow in the infants or in born because it is a non surgical method and cause no pain, no blood loss [1]. This method has advantage over the measurement of the lung volume changes and estimation of stroke volume changes in the aged peoples. The advantage is due to non surgical method which will not cause any strain for the aged people. This is used to indicate the presence or the absence of venous thrombosis. It is an alternative to the venography method [2].

This method is very convenient and easy to use. The inconvenience caused to the people due to the venography method has been overcome in this method of impedance plethysmography. There is no any special dye is injected into the bone marrow or veins. The dye has to be injected constantly via a catheter [17, 18]. Therefore a venography is an invasive procedure. Normally the catheter is entered by the groin and the doctor moves it to the required place navigating through the vascular system. This method innovatively found a solution for those problems.

2 Review of Existing Methods

The problem addressed in this communication concerns the availability of an appropriate instrument for non-invasive measurement of blood flow in human limbs [3, 4]. The most common method used for determination of blood flow in humans is venous occlusion plethysmography, a technique used by physiologists for close to a century. This method is an invasive method herein a dye is being introduced into the veins. This dye is tracked and blood flow is being monitored. Modern venography methods use electronic devices and monitors for the tracking of blood flow. Early plethysmography consisted of either or air-filled compartments in which the limb segment was sealed. These devices also have their respective advantages and disadvantages. Changes in air volume are influenced by body temperature, therefore requiring a correction factor to enable accurate volume determination. The accuracy in blood flow measurements yielded by water plethysmographs and strain gauge plethysmographs is, however, not quantitatively different Because air-filled plethysmographs may also be used to examine the influence of temperature perturbation on blood flow, we chose to explore this method in the present communication. Air-filled plethysmographs were traditionally constructed from glass or brass and other metals to form either conical or boxed-shaped designs [5]. By use of these materials, some of which are no longer appropriate, the Air-filled plethysmograph would prove difficult to manufacture in great numbers.

2.1 Constraints in the Existing Methods

The venographic method is an invasive method wherein it is inappropriate to be used for blood flow measurements among infants, aged people and weaker patients as this method needs to tear the vein. Another major disadvantage is that this method needs good electronic infrastructure for and this adds up to the complexity. In the air filled plethysmography changes in air volume are influenced by body temperature, therefore requiring a correction factor to enable accurate volume determination [19, 20]. Hence the accuracy is a major constrain here, but often we end up in circumstances where we need precise values. Also the inclusion of many parameters makes this method more complex. Application of finger plethysmography requires some limiting considerations [6]. Firstly, PWA is determined by a number of hemodynamic factors including arterial inflow, venous outflow, cardiac stroke volume, venous return to the heart as well as alterations of autonomic neural control. Moreover, the position of the finger relative to the heart level, arm and hand movements (e.g. in sleep studies after arousal), and pre-constriction of finger arteries (e.g. low surrounding temperature, excitement, and stress), may affect the signal [6, 7]. Finally, the agreement between baseline digital blood flow and PWA has been reported to vary to a great extent between Subjects. Therefore, only within-subject changes in PWA during a limited time interval were evaluated in the current study In order to overcome these problems we developed a new methodology called Impedance Plethysmography—Blood Flow Measurement at Foot.

3 Impedance Plethysmography

Impedance plethysmography, also called impedance test or blood flow or impedance phlebography, is a non-invasive test that uses electrical monitoring in the form of resistance (impedance) changes to measure blood flow in veins of the leg [5]. Information from this test helps doctors detect deep vein thrombosis (blood clots or Thrombophlebitis) [18, 20].

- Detect blood clots lodged in the deep veins of the leg.
- Screen patients who are likely to have blood clots in the leg.
- Detect the source of blood clots in the lungs (pulmonary emboli)

Accurate diagnosis of deep vein thrombosis (DVT) is critical because blood clots in the legs can lead to more serious problems. If a clot breaks loose from a leg vein, it may travel to the lungs and lodge in a blood vessels in the lungs. Blood clots are more likely to occur in people who have recently had leg injuries, surgery, cancer, or a long period of bed rest. Because this test is non-invasive, it can be done on all patients and is easy to perform. However, the accuracy of the results is affected if the patient does not breathe normally or keep the leg muscles relaxed [8]. Compression of the veins because of pelvic tumors or decreased blood flow, due to shock or any condition that reduces the amount of blood the heart pumps, may also change the test results. Both false-positives (e.g. when thrombi are non-occlusive) and false negatives have been reported using this technique, which justifies over a period of 7–10 days for patients with initial negative results. Success rates for this test have been estimated at anywhere from 65–66 % to 92–98 %. Using conductive jelly, the examiner strategically places four electrodes on the patient's calf (the four-electrode configuration yields a more uniform and precise current density and consequent measurement result [9, 10]. These electrodes are connected to an instrument called a plethysmography, which records the changes in electrical resistance that occur during the test and produces a graph of the results. The patient must lie down and raise one leg at 30° angle so that calf is above the level of the heart. The examiner then wraps a pressure cuff around the patient's thigh and inflates it to a pressure of 45–60 cm of water for 45 s [11, 12]. The plethysmograph records the electrical impedance that corresponds to change in the volume of blood in the vein at the time the pressure is exerted and again three seconds after the cuff is deflated. This procedure is repeated several times in both legs. This test takes 30–45 min, costs an estimated $50–$100, and results can be available within a few minutes. Impedance plethysmography works by measuring the resistance to the transmission of electrical energy (impedance) [18]. This resistance is dependent upon the volume of blood flowing through the veins [13, 16]. By graphing the impedance, the doctor or technician can tell whether a clot is obstructing blood flow. In the first stage the voltage is measured with the help of an impedance circuitry and later it is replaced with electrodes and the voltage is measured. Patients undergoing this test do not need to alter their diet, change their normal activities, or stop taking any medications [20, 21]. The patient may resume normal or postoperative activities after the test.

4 Block Diagram

4.1 The Non-Inverting Op Amp

The non-inverting op amp has the input signal connected to its non inverting input, thus its input source sees infinite impedance. There is no input offset voltage because VOS = VE = 0, hence the negative input must be at the same voltage as the positive input. The op amp output drives current into RF until the negative input is at the voltage, VIN. This action causes VIN to appear across RG.

The voltage divider rule is used to calculate VIN; VOUT is the input to the voltage divider, and VIN is the output of the voltage divider. Equation (1) is written with the aid of the voltage divider rule, and algebraic manipulation yields Eq. (2) in the form of a gain parameter. When RG becomes very large with respect RF, (RF/RG) $\Rightarrow$ 0 Eq. (1) reduces to Eq. (2).

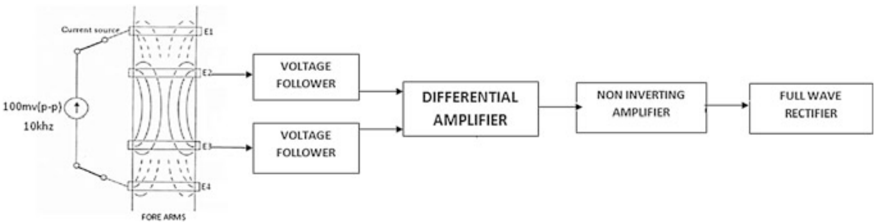


Fig. 1 Block diagram of impedance plethysmography

$$V_{in} = V_{out} \frac{R_C}{R_C + R_F} \tag{1}$$

$$\frac{V_{out}}{V_{in}} = \frac{R_C + R_F}{R_C} = 1 + \frac{R_F}{R_C} \tag{2}$$

Under these conditions VOUT = 1 and the circuit becomes a unity gain buffer. RG is usually deleted to achieve the same results, and when RG is deleted, RF can also be deleted (RF must be shorted when it is deleted). When RF and RG are deleted, the op amp output is connected to its inverting input with a wire. Some op amps are self-destructive when RF is left out of the circuit, so RF is used in many buffer designs. When RF is included in a buffer circuit, its function is to protect the inverting input from an over voltage to limit the current through the input ESD (electro-static discharge) structure (typically <1 mA), and it can have almost any value (20 k is often used). RF can never be left out of the circuit in a current feedback amplifier design because RF determines stability in current feedback amplifiers. Notice that the gain is only a function of the feedback and gain resistors; therefore the feedback has accomplished its function of making the gain independent of the op amp parameters. The gain is adjusted by varying the ratio of the resistors. The actual resistor values are determined by the impedance levels that the designer wants to establish. If RF = 10 k and RG = 10 k the gain is two as shown in Eq. 2, and if RF = 100 k and RG = 100 k the gain is still two. The impedance levels of 10 k or 100 k determine the current drain, the effect of stray capacitance, and a few other points. The impedance level does not set the gain; the ratio of RF/RG does.

4.2 Voltage Follower with Differential Amplifier

The differential amplifier circuit amplifies the difference between signals applied to the inputs. Superposition is used to calculate the output voltage resulting from each input voltage, and then the two output voltages are added to arrive at the final output voltage. Op amp input voltage resulting from the input source, V_1, is calculated in Equations. The voltage divider rule is used to calculate the voltage, V_+, and the non inverting gain equation is used to calculate the non inverting output voltage, $V_{OUT}1$.

$$V_+ = V_1 \frac{R_2}{R_1 + R_2} \tag{3}$$

$$V_{out1} = V_+(G_+) = V_1 \frac{R_2}{R_1 + R_2} \left(\frac{R_3 + R_4}{R_3} \right) \tag{4}$$

The inverting gain equation is used to calculate the stage gain for VOUT2 in Equation. These inverting and non inverting gains are added in Equation.

$$V_{out2} = V_2\left(\frac{R_4}{R_3}\right) \tag{5}$$

$$V_{out} = V_1 \frac{R_2}{R_1 + R_2}\left(\frac{R_3 + R_4}{R_3}\right) - V_2\left(\frac{R_4}{R_3}\right) \tag{6}$$

It is now obvious that the differential signal, (V_1–V_2), is multiplied by the stage gain, so the name differential amplifier suits the circuit. Because it only amplifies the differential portion of the input signal, it rejects the common-mode portion of the input signal. A common-mode signal is illustrated in figure. Because the differential amplifier strips off or rejects the common-mode signal, this circuit configuration is often employed to strip dc or injected common-mode noise off a signal. The disadvantage of this circuit is that the two input impedances cannot be matched when it functions as a differential amplifier, thus there are two and three op amp versions of this circuit specially designed for high performance applications requiring matched input impedances.

4.3 The Full Wave Rectifier

The first building block in the dc power supply is the full wave rectifier. The purpose of the full wave rectifier (FWR) is to create a rectified ac output from a sinusoidal ac input signal. It does this by using the nonlinear conductivity characteristics of diodes to direct the path of the current. If we consider a simple, piece-wise linear model for the diode IV curve, the diode forward current is zero until Bias > = $V_{threshold}$, where $V_{threshold}$ is 0.6–0.8 V. The current increases abruptly as V_{bias} increases further. Due to this turn-on or threshold Voltage associated with the diode in forward bias, we should expect a 0.6–0.8 V voltage drop across each forward biased diode in the rectifier bridge. In the case of the full wave rectifier diode bridge, there are two forward biased diodes in series with the load in each half cycle of the input signal. The maximum output voltage (across load) will be V_{in} –2 $V_{threshold}$, or ~ Vin—1.4 V. Since some current does flow for voltage bias below $V_{threshold}$ and the current rise around is $V_{threshold}$ is more gradual than the piece-wise model, the actual diode performance will differ from the simple model. In reverse bias (and neglecting reverse voltage breakdown), the current through the diode is approximately the reverse saturation current, Io. The voltage across the load during reverse bias will be V_{out} = Io R_{load}. In specifying a diode for use in a circuit, you must take care that the limits for forward and reverse voltage and current are not exceeded.

4.4 Hardware Descriptions

The potential of the Ag/Agcl electrode is determined by the silver-ion activity in solution, which is inversely related to chloride activity through the solubility product. Any additional ion present in sufficient concentration to form an insoluble silver salt, such as may be found in solutions of biological origin, will contribute to the resultant potential. The four electrodes (Ag/Agcl) are placed on the surface of the body (either on the surface of foot or hand). In Fig. 1 the first and fourth electrode is connected to function generator which acts as the current source. The function generator is set with 10 KHZ frequency and 100 mill volt peak to peak. From the second and third electrode the input is generated and it is given to the voltage follower. Then the signal is amplified followed by Differential amplifier, Non Inverting amplifier and Full Wave Rectifier. Finally the output voltage is measured in CRO.

5 Results and Discussion

Normally, inflating the pressure cuff will cause a sharp rise in the pressure in the calf because blood flow is blocked. When the cuff is released, the pressure decreased rapidly as the blood flows away. If a clot is present; the pressure in the calf veins will already be high. It does not become sharply higher when the pressure cuff is tightened. When the pressure cuff is deflated, the clot blocks the flow of blood out of the calf vein. The decrease in pressure is not as rapid as when no clot is present and the shape of the resulting graph is different, all of which is indicative of obstruction of major deep veins. Table 1 shows the values obtained through various impedance values and its calculated value. The blood flow in millilitre is also calculated using Eq. 7.

$$\Delta v = \frac{\rho L^z}{Z^2} \Delta Z \tag{7}$$

Figure 2 shows the Response of Impedance Meter for various simulated Resistance Values. This Graph shows the nonlinearity in the resistance region of 30–100 ohms. based on the above tabulated readings blood flow may be a turbulent one than a linear nature for standard atmospheric pressure.

Figure 3 is depicted the error Plot for Simulated and Measured Resistance. From the figure 3 it is also observed that more prominent error is in the measurable flow region. Hence an non linear artificial intelligence based method is to predict the blood flow in the peripheral region of the body. The Impedance plethysmography method involves the measurement of the electrical parameters of the body and then this value is directly transduced into an equivalent desired value as per our requirements thus the accuracy of the measurement is high. Using water as a medium for displacement lends itself to studies concerned with temperature perturbation that are more difficult to conduct using mercury-in-Silastic strain

Table 1 Analysis of impedance meter error

Impedance (ohms)	Calculated amplifier O/P V_1(Volts)	Measured value O/P V_2(Volts)	% of error $((V_1-V_2)/V_1)$x100	Blood flow ΔV ml
10	0.221	0.23	−4.07	0.54
13	0.31	0.30	3.22	0.702
15	0.442	0.34	23.07	0.81
20	0.4862	0.45	7.45	1.08
22	0.553	0.50	9.58	1.18
25	0.663	0.56	15.5	1.35
30	0.729	0.66	12.9	1.62
33	0.862	0.732	15.08	1.782
39	0.862	0.866	−0.46	2.106
44	0.972	0.98	−0.82	2.376
56	1.23	1.28	−4.06	3.024
69	1.518	1.50	1.18	3.726
76	1.672	1.68	−0.47	4.104
88	1.96	1.98	−1.02	4.752
99	2.17	2.19	−0.92	5.346
100	2.27	2.31	−1.76	5.4
120	2.39	2.45	−2.51	6.48
156	2.52	2.56	−1.58	8.424
160	2.63	2.69	−2.28	8.64
220	2.76	2.29	−1.08	11.88

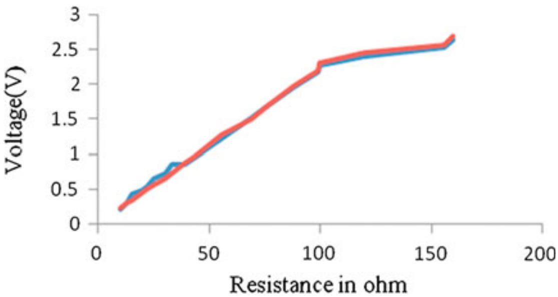


Fig. 2 Response of impedance meter for various simulated resistance values

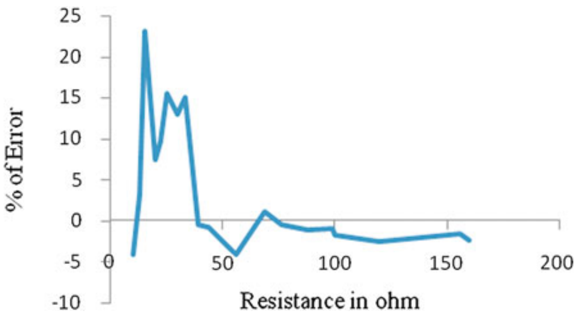


Fig. 3 Error plot for simulated and measured resistance

gauge plethysmography or air plethysmography. One example of such a study is to examine the influence of local temperature changes on predominantly skin blood flow in different body positions. The technique of limb plethysmography has many more applications for both research and teaching. The instrument proves to be robust, accurate, and inexpensive to manufacture in numbers with the assistance of a standard University workshop. Students evaluated use of the plethysmography positively as an aid to learning. There are limitations to all experimental techniques and the one presented here is no exception. The main limitation of our simple plethysmograph design is lack of a mechanism to control water temperature, which is known to influence forearm blood flow measurement. Other investigators have built plethysmography complete with an external water bath into which cold or warm water is added without affecting the volume of water in the plethysmograph proper. This design would, however, complicate the construction procedure tremendously and make it difficult to produce the device in any great numbers.

6 Conclusion

Measurement of compartmental blood flow by finger plethysmography provides a tool for continuous non-invasive assessment of changes in digital blood flow. This project is an alternative way to measure the blood flows in a very accurate manner and it is less expensive. This method will be robust and accurate. This simple plethysmograph design is an alternative to other commercially available equipments. Thus we made a hardware design by placing the electrodes (Ag/Agcl) on our body surface (either foot or hand surface) and the change in output voltage is checked with various persons. It gets varied according to the age of person and it is determined on the basis of rate of blood flow. The variations in the blood flow can be analysed based on various activities such as (walking, running, sleeping etc.). Finally the value of change in voltage is converted to blood flow in ml with respect to length of the vein (e.g. 3 ml for 5 cm). Thus the results are analysed from hardware design implementation. In future, this implementation is to be interfaced in LABVIEW software which can be used in real time applications.

References

1. Barcroft H, Edholm OG (1943) The effect of temperature on blood flow and deep temperature in the human Forearm. J Physiol 102:5–20
2. Comroe JH, Botelho SY, DuBois AB (1959) Design of a body plethysmograph for studying cardiopulmonary physiology. J Appl Physiol 14:439–444
3. Christ F, Gamble J, Baschnegger H, Gartside IB (1997) Relationship between venous pressure and tissue volume during venous congestion plethysmography in man. J Physiol 503(2):463–467

4. Fleckenstein KJ (1984) The Mosso plethysmograph in 19th- century physiology. Med Instrum 18:330–331
5. Gibbs G, Gregory R, Moore I (1997) Teaching more students 7: labs and practicals with more students and fewer resources. Oxford Centre for Staff Development, Oxonian Bewley, Oxford, pp 8–32
6. Halliwill JR, Minson CT, Joyner MJ (1999) Measurement of limb venous compliance in humans: technical considerations and physiological findings. J Appl Physiol 87:1555–1563
7. Kramer K, Lochner W, Wetterer E (1965) Methods of measuring blood flow. In: Handbook of physiology. Circulation. American Physiological Society, Bethesda, sect 2, vol II, chap 38, pp 1277–1290
8. Bos WJ, van den Meiracker AH, Wesseling KH, Schalekamp MA (1995) Effect of regional and systemic changes in vasomotor tone on finger pressure amplification. Hypertension 26:315–320
9. Burch GE (1965) Digital plethysmography. Grune & Stratton Inc, New York
10. Burton AC (1939) The range and variability of the blood flow in the human fingers and the vasomotor regulation of body temperature. Am J Physiol 127:437–453
11. Jindal GD, Nerukkar SN, Pendukar SA, Babu JP, Kelkar MD, Despande AK, Parulkar GB (1990) Diagnosis of peripheral arterial occlusive disease using impedance plethysmography. J Postgrad Med 36:147–153
12. Nyober J (1980) Regional pulse volume and perfusion flow measurements: electrical impednce plethysmography. Arch Int Med 105:264–276
13. Pethig R (1987) Dielectric properties of live tissues-clinical physics and physiological measurement. J Clin phy Physiol Meas 8:5–12
14. Joyner MJ, Dietz NM, Shepherd JT (2001) From Belfast to Mayo and beyond: the use and future of plethysmography to study blood flow in human limbs. J Appl Physiol 91:2431–2441
15. Burch GE (1986) Influence of sublingual nitroglycerin on the digital circulation of man. Angiology 37:809
16. Mc Neil MF, Thro E (1994) Fuzzy logic a practical approach forwarded by R. Vage Ap Professional
17. Hu YH, Palreddy S, Tompkins J (1997) A patient adaptive ECG beat classifier using a mixture of experts approach. IEEE Trans Biomed Eng 44(9):891–900
18. Faezipour M, Tiwari TM, Saeed A, Nourani M, Tamil LS (2009) Wavelet-based denoising and beat detection of ECG signal. In: Proceedings IEEE-NIH life science system application workshop, pp 100–103
19. Zhang D (2005) Wavelet approach for ECG baseline wander correction and noise reduction. In: Proceedings of the 27th IEEE annual conference engineering in medicine and biology, pp 1212–1215
20. de Chazal P, O'Dwyer M, Reilly RB (2004) Automatic classification of heartbeats using ECG morphology and heartbeat interval features. IEEE Trans Biomed Eng 51(7):1196–1206
21. Zarychta P, Smith FE, King ST, Haigh AJ, Klinge A, Zheng D, Stevens S, Allen J, Okelarin A, Langley P, Murray A (2007) Body surface potential mapping for detection of myocardial infarct sites. In: Proceedings IEEE computers in cardiology, pp 181–184

A Brightness Preserving Contrast Enhancement Method Based on Clipped Histogram Equalization

C. S. Sarada and M. Wilscy

Abstract A good brightness preserving contrast enhancement method has several applications especially in consumer electronics. A novel contrast enhancement method based on histogram equalization has been proposed in this paper. The method divides the histogram of an image into four sub histograms. Then, clipping is applied on each sub histogram. The clipping threshold of each sub histogram is taken as the difference between the median and standard deviation of the occupied intensities which gives a smoother cumulative distribution function and leads to better equalization of the sub histograms resulting in a good enhancement. Each sub histogram is assigned a new range and histogram equalization is done independently.

Keywords Contrast enhancement · Brightness preserving · Histogram equalization · Standard deviation

1 Introduction

Contrast generally refers to the difference in luminance or gray level values in an image. It is the property which enables to differentiate the varying details in an image. In general the objective of contrast enhancement methods is to improve the

C. S. Sarada (✉) · M. Wilscy
Department of Computer Science, University of Kerala, Kariavattom Campus, Thiruvananthapuram, India
e-mail: sarada.cs@gmail.com

M. Wilscy
e-mail: wilsyphilipose@hotmail.com

Mohan S. and S. S. Kumar (eds.), *Proceedings of the Fourth International Conference on Signal and Image Processing 2012 (ICSIP 2012)*, Lecture Notes in Electrical Engineering 222,
DOI: 10.1007/978-81-322-1000-9_5,

contrast by preserving the natural look of the images, maintaining brightness as such, avoiding noise amplification and unwanted artifacts. Contrast enhancement has several applications in areas like consumer electronics, medical image processing and satellite image processing.

Among the several techniques available for contrast enhancement of an image, histogram equalization is one of the most conventional ones. In the conventional Histogram Equalization (HE) [1] method, the intensity values are spread out and redistributed to obtain a uniformly distributed flat output histogram. The cumulative distribution function of the input histogram acts as a mapping function to find the new intensity levels corresponding to the old ones. The major drawback of the HE methods lies in the fact that irrespective of the input mean brightness, the mean brightness of the output images is obtained as the middle gray level, thus it fails in preserving the input mean brightness of an image. The output of HE also suffers intensity saturation and over enhancement when quantum jump occurs in the cumulative distribution function of the histogram [2].

A new method known as Brightness preserving Bi-histogram equalization (BBHE) [3] was proposed by Kim in 1997. BBHE divides the input histogram into two sub histograms based on the mean of the input image. Then HE is performed on the two sub histograms separately. Even though the method attempts to preserve the input mean brightness, it still results in saturation in some cases.

Later, Dualistic sub image histogram equalization DSIHE [4] was proposed by Wan et al. in 1999. Here the input histogram is divided into two sub histograms in such a way that both the halves contain equal number of pixels. It preserves mean brightness and also increases the entropy of the image. But there are several cases where DSIHE also fails to avoid unwanted artifacts.

Chen and Ramli proposed another method called minimum mean brightness error bi-histogram equalization (MMBEBHE) [5] in 2003. This method divides the input histogram at the optimal point where the mean brightness error between input and enhanced images is minimum. Even though the mean brightness is preserved the computational complexity is significantly high. Chen and Ramli later proposed Recursive mean-separate histogram equalization (RMSHE) [6] in 2003. Here segmentation based on mean is performed recursively to obtain 2^r sub histograms. Then HE is applied on each sub histogram separately. Sim et al. proposed RSIHE [7] in 2007. In RSIHE recursive segmentation is based on median. Even though RMSHE and RSIHE serve in preserving the brightness they do not significantly improve the contrast and also lead to unwanted artifacts.

Kim and Chung proposed Recursively separated and weighted histogram equalization (RSWHE) [8] in 2008. The input histogram is segmented recursively based on mean (RWSHE-M) or median (RWSHE-D) of the image. Then, the sub histograms are modified by applying a weighting process based on normalized power law function. HE is applied on the modified sub histograms separately. Ooi et al. proposed Bi-histogram equalization with a plateau limit (BHEPL) in 2009. BHEPL [9] is a hybrid of BBHE and clipped histogram equalization (CHE) [10]. The input histogram is divided into two sub histograms based on the mean brightness of the input image as in BBHE. Then a clipping process is performed on each sub histogram. Finally HE is applied on the sub histograms separately.

Later in 2010 Ooi and Isa proposed BHEPLD and DQHEPL methods [11]. BHEPLD is a modification of BHEPL where the clipping threshold is the median of the occupied intensity instead of mean. DQHEPL is an extension of RSIHE. Here 'r' is fixed as two. Clipping is performed on each sub histogram by finding the thresholds separately. Here the threshold is the average of the occupied intensity.

Clipping helps in avoiding quantum jump caused due to steep change in the cumulative distribution function (cdf) which leads to over enhancement. A change in the clipping threshold results in significant improvement in the output. Determining the correct clipping threshold plays a vital role in obtaining excellent output without any unwanted artifacts. A good brightness preserving contrast enhancement method which is devoid of noise amplification, intensity saturation and over-enhancement is very essential for most of the applications.

A variant of DQHEPL method is proposed in this work in which the clipping threshold for each sub histogram is set as the difference between median and standard deviation of the occupied intensities. The standard deviation represents the net variation in a histogram/sub histogram. The clipping threshold value calculated based on the difference between median and standard deviation leads to a cdf with smooth variation (no steep changes) which on equalization gives uniform spread across the range resulting in a well enhanced image. The cdf obtained using the proposed method after clipping is smoother than the cdf obtained in DQHEPL which in turn results in a better output. The proposed method is tested with numerous images and is found to be successful in preserving brightness better than the existing methods while maintaining the original shape of the histogram. It also has a better AAMBE and Average entropy value which is proved based on several test images.

This paper has four sections. The Sect. 2 will elaborate on the proposed DQHESTD method. Section 3 is a detailed discussion on results obtained. The results using the existing methods are compared in detail. The conclusions are presented in Sect. 4.

2 Dynamic Quadrants Histogram Equalization Using Median and Standard Deviation for Clipping (DQHESTD)

In this section we discuss the proposed method in detail. This method has four steps which are histogram segmentation, histogram clipping, new gray level allocation and histogram equalization as shown in Fig. 1 [11].

2.1 Histogram Segmentation

The histogram is segmented into four sub histograms in such a way that each portion contains equal number of pixels. m_0 and m_4 are the minimum and maximum intensity values of the input image respectively.

$$m_1 = 0.25 \times \{I_{width} \times I_{height}\} \tag{1}$$

$$m_2 = 0.5 \times \{I_{width} \times I_{height}\} \tag{2}$$

$$m_3 = 0.75 \times \{I_{width} \times I_{height}\} \tag{3}$$

where m_1, m_2 and m_3 are the intensity values with number of pixels equal to 0.25, 0.5 and 0.75 of the total number of pixels in the histogram of the image. I_{width} and I_{height} represent the size of the image.

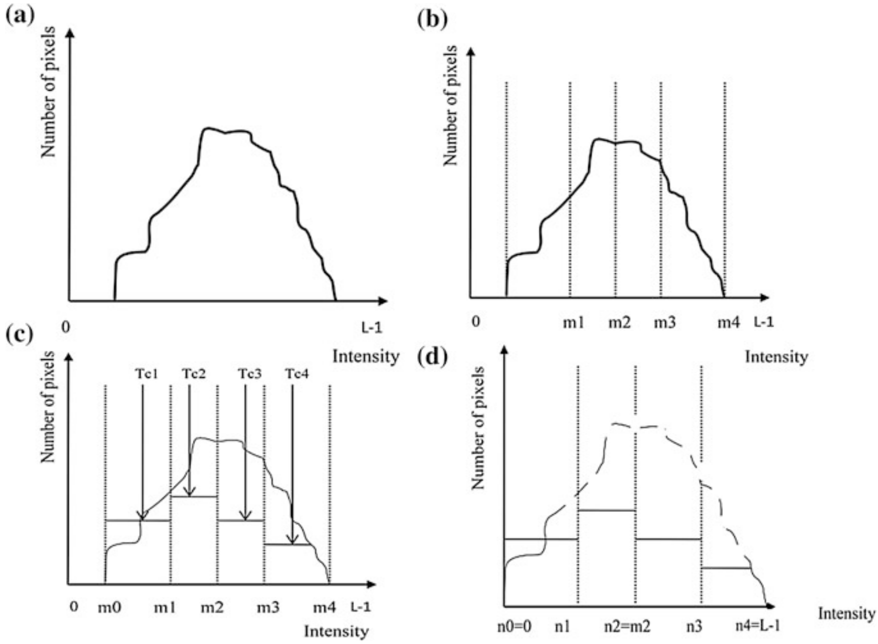


Fig. 1 **a** Original histogram. **b** Histogram segmentation. **c** Histogram clipping. **d** New gray level allocation

2.2 Histogram Clipping

The clipping threshold of each sub histogram is the difference between the median and standard deviation of the occupied intensities. It is represented as Tc_i.

$$Tc_i = median(h_i(X_k)) - \alpha * standarddeviation(h_i(X_k)) \tag{4}$$

$h_i(X_k)$ represents the number of pixels at intensity level k in the ith subhistogram where $k = m_{i-1}\ to\ m_i$. The clipping threshold for each sub histogram has to be calculated. As there are 4 sub histograms i = 1, 2, 3, 4.

In this clipping method the threshold might become zero or less than zero in some cases. This will lead to loss of detail. α is a factor which is used to optimize the clipping threshold. The value of α lies in the range [0, 1]. A judicious selection of this parameter is critical in obtaining a smooth cdf. If $\alpha = 1$ results in a zero or negative Tc, then it takes the next smaller value in the range which is a multiple of 0.25 and this process is continued till we get a non-zero non-negative threshold.

$$h_{Ci}(X_k) = \begin{cases} h_i(X_k) & h_i(X_k) \le Tc_i \\ Tc_i & h_i(X_k) > Tc_i \end{cases} \tag{5}$$

where $h_{ci}(X_k)$ is the clipped histogram at intensity level k in the i^{th} sub histogram. Any intensity k with number of pixels $h_i(X_k)$ greater than the threshold is clipped and the number of pixels is set to Tc_i as in Eq. 5.

2.3 New Gray Level Allocation

This is done to balance the enhancement space for each sub histogram which makes the final histogram span the full range. The new gray levels are represented as n_0, n_1, n_2, n_3 and n_4.

n_0 and n_4 are selected as the lowest and highest gray levels available ($n_0 = 0$ and $n_4 = L - 1$).

$$n_1 \cong m_2 * \frac{m_1 - m_0}{m_2 - m_0}. \tag{6}$$

$$n_2 \cong m_2 \tag{7}$$

This gray level (m2) represents the mean brightness of the input and thus is not changed while allocating the new gray level which ensures brightness preserving.

$$n_3 \cong (L - 1 - m_2) * \frac{m_3 - m_2}{m_4 - m_2} + m_2. \tag{8}$$

2.4 Histogram Equalization

Each sub histogram is equalized separately. The transform function for each pixel intensity X is:

$$Y(X) = n_{i-1} + (n_i - n_{i-1}) * \frac{\sum_{k=m_{i-1}}^{m_i} h_{ci}(X_k)}{M_i}. \tag{9}$$

where Y represents the new pixel intensity corresponding to X

$$M_i = \sum_{k=m_{i-1}}^{m_i} h_{ci}(X_k) \tag{10}$$

where M_i is the total number of pixels in the *i*th clipped sub histogram.

3 Results and Discussion

The proposed method is tested on numerous images. The comparison with some of the state of the art methods for contrast enhancement and the results obtained is discussed in detail. The methods used for comparison are HE, BBHE, BHEPLD and DQHEPL. The images of *'Einstein', 'Airplane' and 'Clock'* are used for analysis. The measures used for quantitative analysis are Average Absolute mean brightness error and Average Entropy. The measures are calculated as described below:

Average Absolute mean brightness error,

$$AAMBE = \frac{1}{N}|B(X) - B(Y)| \tag{11}$$

where B(X) is the mean brightness of the input image, B(Y) mean brightness of the output image and N is the number of test images used.

AMBE is the difference in the mean brightness of the input and output images. A good brightness preserving method will have a low AMBE.

Absolute mean brightness error,

$$AMBE = |B(X) - B(Y)| \tag{12}$$

If there are 10 test images then N = 10. AMBE of these 10 images are used to calculate AAMBE.

AverageEntropy,

$$AE = \frac{1}{N}\sum_{k=0}^{255} P_{out}(X_k) * log_2 P_{out}(X_k) \tag{13}$$

N is the number of test images used. AE is the average entropy of 'N' test images.

Entropy is the average information content in an image. Higher the entropy of the output better is the method.

$$Entropy = \sum_{k=0}^{255} P_{out}(X_k) * log_2 P_{out}(X_k) \quad (14)$$

The qualitative analysis on a few of the test images is shown in the figures below. Using HE and BBHE it is observed in Fig. 2b, c that the face shows intensity saturation. The color of the coat which is actually gray looks darker. The output of BHEPLD Fig. 2d is acceptable but certain artifacts are seen on the coat, the forehead is also over enhanced. DQHEPL, Fig. 2e has intensity saturation on the forehead and cheeks. The proposed method shown in Fig. 2f enhanced the image without any unwanted artifacts. The color of the coat is preserved and there is no intensity saturation on the face. The histogram analysis of this image has been shown in Fig. 3. Using HE, Fig. 3a, the shape of the original histogram is not at all preserved. Fig. 3c–e the shape of the original histogram is preserved to an extent and it is stretched to the full range but there is intensity saturation at the right end. The histogram obtained after applying the proposed method Fig. 3f, preserves the original shape exactly. None of the other methods maintain the gradual rise at the leftmost end of the histogram. The rounded portion in Fig. 3a, f

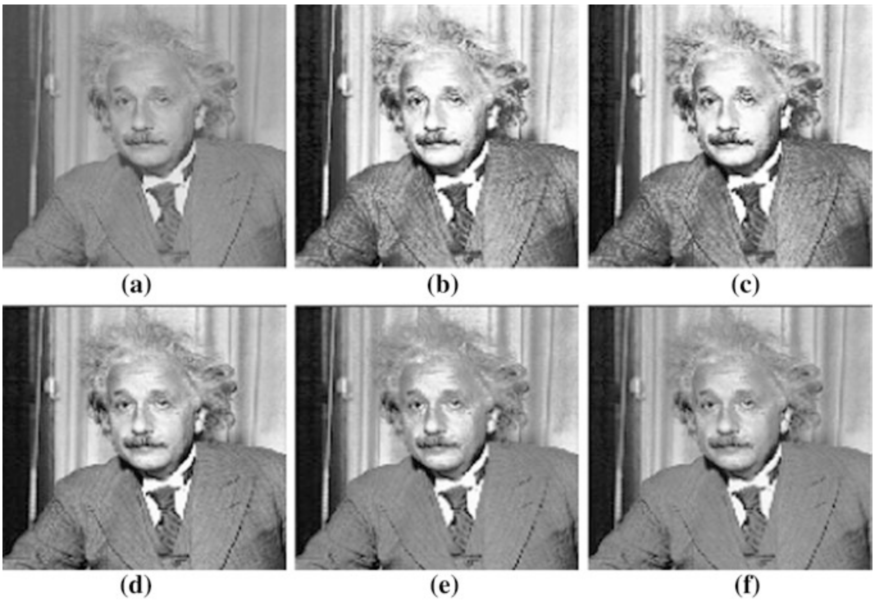


Fig. 2 Einstein. **a** Original. **b** HE. **c** BBHE. **d** BHEPLD. **e** DQHEPL. **f** Proposed method

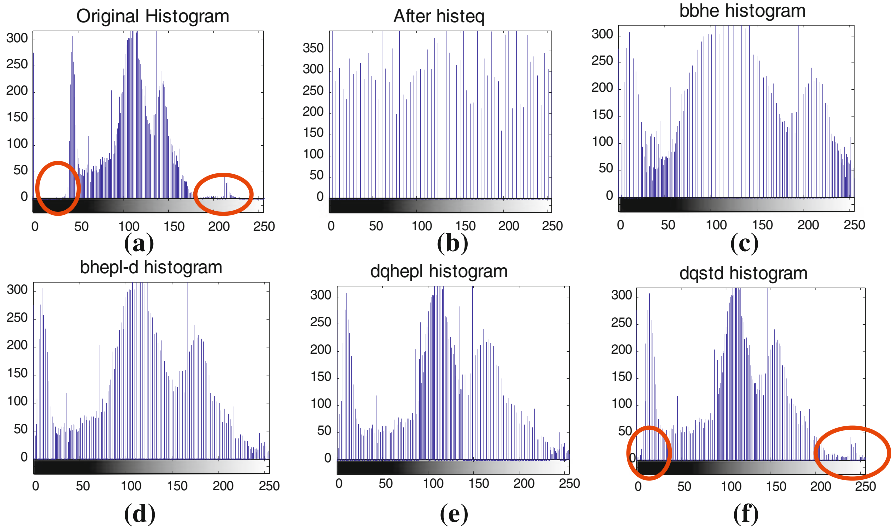


Fig. 3 Histogram analysis of 'Einstein'. **a** Original. **b** HE. **c** BBHE. **d** BHEPLD

clearly shows this. The intensity saturation is also reduced using the proposed method.

From Fig. 4 it is seen that the output of HE, Fig. 4b shows over enhancement with dark artifacts on the hill and sky regions. The output of BBHE, BHEPLD and DQHEPL Fig. 4c–e are acceptable but the hill region is slightly saturated. The proposed method reduces the intensity saturation to an appreciable extent. It preserves the brightness and increases the entropy better than other methods. The histogram analysis is shown in Fig. 5. Even though HE stretches the histogram; it does not preserve the shape. The intensity saturation towards the darker side of histogram in methods BBHE, BHEPLD and DQHEPL is visible in Fig. 5c–e.The histogram of the proposed method Fig. 5f reduces the intensity saturation problem and preserves the original shape of the histogram.

The result analysis of 'Clock' image is shown in Fig. 6. Using HE, Fig. 6b the enhanced image looks unnatural and fails to enhance features of the book and clock. The output of BBHE and DQHEPL Fig. 6c, e is acceptable but has slight line artifacts on the wall behind the clock. BHEPLD Fig. 6d gives good enhancement without any unwanted artifacts but the wall is slightly brighter than the original. The proposed method Fig. 6f enhances the image without producing any unwanted artifacts and preserves the brightness better than the other methods. The histogram of BBHE, BHEPLD and DQHEPL Fig. 7c–e, show slight intensity saturation towards the right end. The proposed method removes the saturation and also preserves the shape of the histogram exactly as the original.

The measurements are taken using 11 standard test images (N = 11). The test set includes the following images:-Einstein, airplane, clock, child, girl in green, couple, image1 [9], image2 [9], tree, child in yellow, board. All the methods used

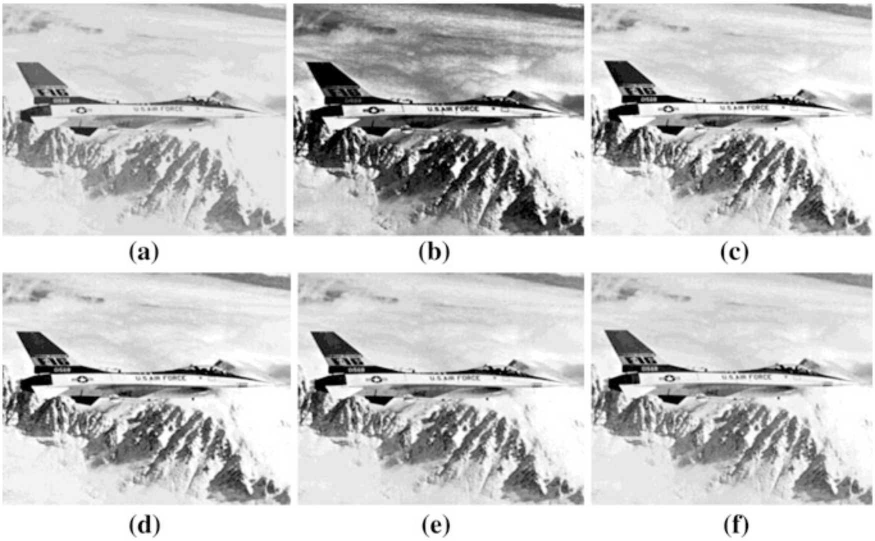


Fig. 4 Airplane. **a** Original. **b** HE. **c** BBHE. **d** BHEPLD. **e** DQHEPL. **f** Proposed method

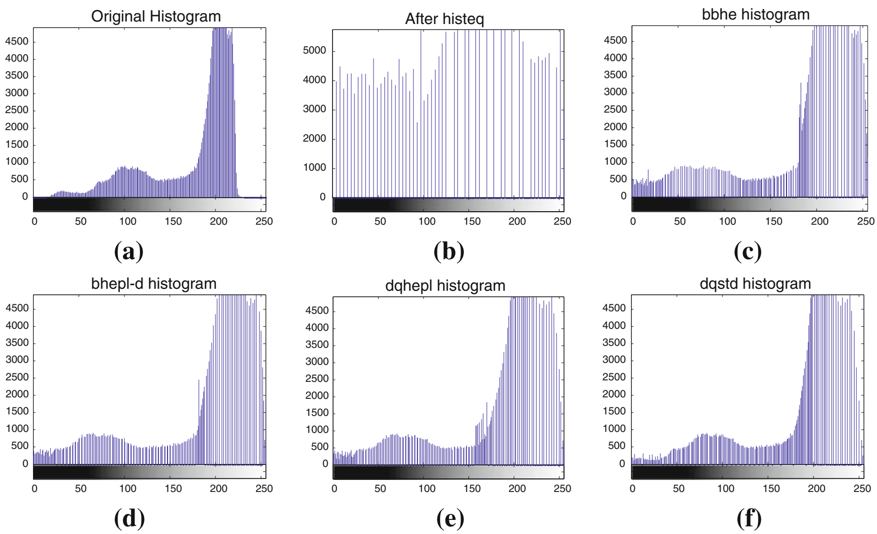


Fig. 5 Histogram analysis of ‘Airplane’. **a** Original. **b** HE. **c** BBHE. **d** BHEPLD. **e** DQHEPL. **f** Proposed method

for comparison were implemented for testing. From Table 1 we can see that HE and BBHE have a very high AAMBE value which shows that it does not preserve brightness. BHEPLD and DQHEPL have good AAMBE and AE values.

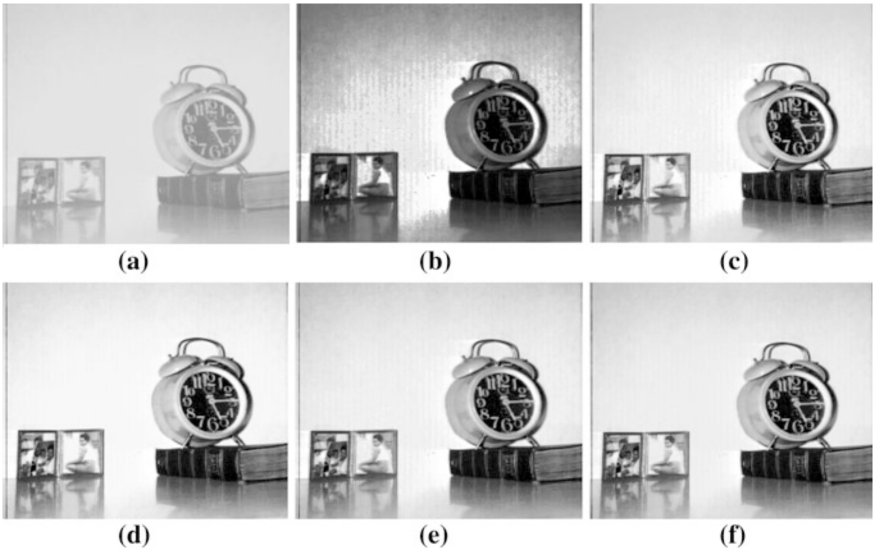


Fig. 6 Clock. **a** Original. **b** HE. **c** BBHE. **d** BHEPLD. **e** DQHEPL. **f** Proposed method

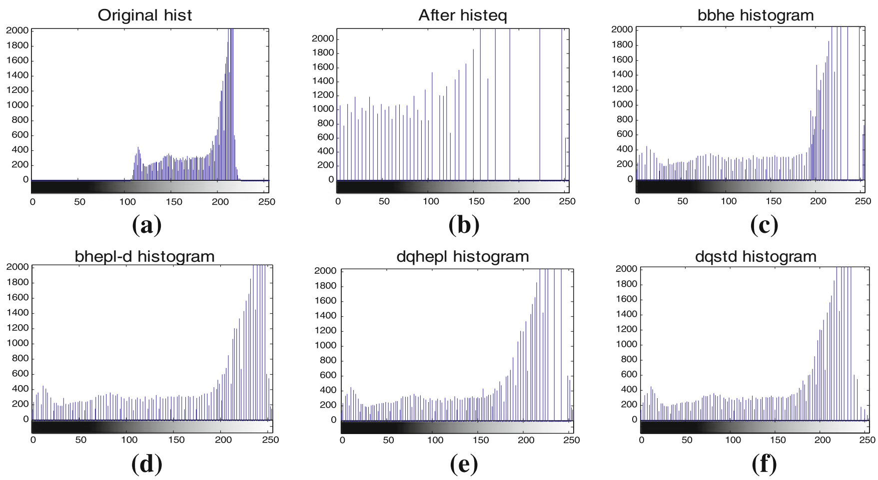


Fig. 7 Histogram analysis of 'Clock'. **a** Original. **b** HE. **c** BBHE. **d** BHEPLD. **e** DQHEPL. **f** Proposed method

The proposed method performs the best with least AAMBE value and highest entropy which shows that it is a good contrast enhancement technique with brightness preserving capability.

Table 1 Quantitative analysis

Method	AAMBE	Average entropy
Original	–	6.54
HE	46.93	5.64
BBHE	17.83	6.38
BHEPLD	8.63	6.50
DQHEPL	7.13	6.48
DQHESTD (proposed)	5.47	6.53

4 Conclusion

A novel contrast enhancement method which preserves brightness has been proposed. The method DQHESTD, is tested on numerous images and the results obtained evidently show that it successfully enhances the image and also exhibits brightness preserving capability. It has been proved quantitatively and qualitatively that the proposed method outperforms the existing methods by enhancing the details in the image without introducing any unwanted artifacts.

References

1. Gonzalez RC, Woods RE (2008) Digital image processing, 3rd edn. Prentice Hall, Upper Saddle River
2. Yun SH, Kim JH, Kim S (2010) Image enhancement using a fusion framework of histogram equalization and Laplacian pyramid. IEEE Trans Consum Electron 56(4):2763–2771
3. Kim YT (1997) Contrast enhancement using brightness preserving bi histogram equalization. IEEE Trans Consum Electron 43(1):1–8
4. Wan Y, Chen Q, Zhang BM (1999) Image enhancement based on equal area dualistic sub-image histogram equalization method. IEEE Trans Consum Electron 45(1):68–75
5. Chen SD, Ramli AR (2003) Minimum mean brightness error bi histogram equalization in contrast enhancement. IEEE Trans Consum Electron 49(4):1310–1319
6. Chen SD, Ramli AR (2003) Contrast enhancement using recursive mean-separate histogram equalization for scalable brightness preservation. IEEE Trans Consum Electron 49(4):1301–1309
7. Sim KS, Tso CP, Tan YY (2007) Recursive sub-image histogram equalization applied to gray-scale images. Pattern Recognit Lett 28:1209–1221
8. Kim M, Chung MG (2008) Recursively separated and weighted histogram equalization for brightness preservation and contrast enhancement. IEEE Trans Consum Electron 54(3): 1389–1397
9. Ooi CH, Kong NSP, Ibrahim H (2009) Bi-histogram equalization with a plateau limit for digital image enhancement. IEEE Trans Consum Electron 55(4):2072–2080
10. Kim T, Paik J (2008) Adaptive contrast enhancement using gain controllable clipped histogram equalization. IEEE Trans Consum Electron 54(4):1803–1810
11. Ooi CH, Isa NAM (2010) Adaptive contrast enhancement methods with brightness preserving. IEEE Trans Consum Electron 56(4):2543–2551

Example Based Super Resolution Using Fuzzy Clustering and Neighbor Embedding

Keerthi A. S. Pillai and M. Wilscy

Abstract A fuzzy clustering based neighbor embedding technique for single image super resolution is presented in this paper. In this method, clustering information for low-resolution (LR) patches is learnt by Fuzzy K-Means clustering. Then by utilize the membership degree of each LR patch, a neighbor embedding technique is employed to estimate high resolution patches corresponding to LR patches. The experimental results show that the proposed method is very flexible and gives good results compared with other methods which use neighbor embedding.

Keywords Fuzzy clustering · Neighbor embedding · Example based super resolution

1 Introduction

In most electronic imaging applications, images with high resolution (HR) are desired and often required [1]. HR means that pixel density within an image is high, and therefore an HR image can offer more details that may be critical in various applications such as medical diagnosis, pattern recognition, remote surveillance, biometric identification etc. The most direct solution to increase spatial

K. A. S. Pillai (✉) · M. Wilscy
Department of Computer Science, University of Kerala, Kariavattom,
Thiruvananthapuram 695581, India
e-mail: keerthias@gmail.com

M. Wilscy
e-mail: wilsyphilipose@hotmail.com

Mohan S. and S. S. Kumar (eds.), *Proceedings of the Fourth International Conference on Signal and Image Processing 2012 (ICSIP 2012)*, Lecture Notes in Electrical Engineering 222,
DOI: 10.1007/978-81-322-1000-9_6,

resolution is either reduce the pixel size (i.e., increase the number of pixels per unit area) by sensor manufacturing techniques or to increase the chip size. However, both methods are limited by the high cost for high precision optics and image sensors [1]. To overcome the limitations of the above approaches, image processing techniques are introduced in super resolution. In almost 30 years, several approaches have been proposed for super resolution [2, 3]. The existing super resolution approaches can be divided into three: based on interpolation, based on degrading model, and example based super resolution. This paper focus on example based super resolution in which a high resolution image is generated from a single low-resolution image, with the help of a set of one or more training images from scenes of the same or different types.

Most of the example based super resolution methods are based on the idea that high-frequency details of low-resolution (LR) can be predicted by learning the co-occurrence relationship between LR patches and their corresponding high-resolution (HR) patches. Based on this idea Freeman et al. [4, 5] proposed an example based super-resolution which uses Markov network for learning the co-occurance relationship between LR patches and corresponding HR patches. But this method is greatly handicapped by its sensitivity to training samples. To overcome the above limitation, Chang et al. [6] proposed a method based on linear embedding which uses the concept of manifold learning [7–10] for HR reconstruction. This method uses the *k* candidate HR patches from the training set and find the optimal weight for each HR patch which minimize the reconstruction error in HR generation. This method has some advantages over the Freeman's approach [5], but the result depends on the neighbor size chosen. Since the edges and neighbor size plays an important role in the HR construction, Chan et al. [11] modified the idea of super-resolution through neighbor embedding and gave more importance to edge detection and feature selection. This method treats the edge and non edge patches independently with different neighbor size while embedding. Since the algorithm depends highly on the edge detection, incorrect edge detection will lead artifacts in the target HR image. To make the LR patches compatible with each other and independent of edge detection, Zhang et al. [12] introduced the concept of clustering in neighbor embedding. They used the Gaussian Mixture Model (GMM) [13, 14] for clustering the training LR patches. This class information is used in the selection of k candidate HR patch in the neighbor embedding.

In CSNE [12], even though soft clustering is used to cluster the training patches, the manner in which the class information used is deterministic. So the selection procedure of candidate training LR patches does not fully utilize the advantage of soft clustering and select the wrong HR patch for reconstruction. Since clustering plays an important role in the selection procedure of neighbour embedding, the incorrect clustering information highly affects the HR reconstruction results. This will affect the target HR image. And also GMM clustering requires large storage and complex processing units due to exponential calculations and a large number of coefficients involved. And the complexity increases exponentially with dimensionality in GMM clustering. So to reduce computational complexity in the training phase we have to choose an alternative method.Taking all these into account we propose an example based super resolution using fuzzy clustering and neighbor embedding algorithm.

The major contribution of our proposed method is as follows.

To cluster the training LR patches fuzzy clustering is used. In the selection process of neighbor patches, the eucledian distance between the input LR patch and training LR patches is modulated by a factor that depends on the fuzzy membership degree obtained after clustering. So that the LR training patches selected in the neighbor embedding stage of the proposed method is more compatible to input LR patch than that in the CSNE [12]. This will improve the visual quality of target HR image. Also the fuzzy clustering reduces the computational complexity in the training phase of clustering. The performance of the proposed method is evaluated quantitatively using PSNR and the proposed method gives better quantitative results than the other methods which use neighbor embedding. The proposed method is able to recover the better visual quality than other algorithms subjectively.

The remainder of the paper is organized as follows: In Sect. 2 the proposed method is described in detail. Experimental results are present in Sect. 3. Section 4 concludes the paper.

2 Proposed Method

In this section first give a brief introduction about fuzzy k means clustering. Then describe the fuzzy clustering based neighbor embedding in detail.

2.1 Fuzzy K-Means Clustering

Numerous problems in real-world applications, such as pattern recognition and computer vision, can be tackled effectively by the fuzzy K-Means algorithms [15–17]. Each pattern is allowed to have memberships in all clusters rather than having a distinct membership to one single cluster. In the proposed method, fuzzy clustering [18, 19] is used to cluster the LR training patches. Each patch is assigned a membership degree in each cluster. The details of clustering are described below.

Let $X = \{X_1, X_2, \ldots, X_n\}$ be a set of n objects in which each object X_i is represented as $\{X_{i1}, X_{i2}, \ldots, X_{im}\}$, where m represents the dimension of the feature vector X_i. To cluster X into k clusters by the fuzzy K-Means algorithm is to minimize the following objective function

$$P(U, Z) = \sum_{j=1}^{k} \sum_{i=1}^{n} \left(u_{i,j}\right)^{\lambda} D_{i,j} \tag{1}$$

subject to

$$\sum_{j=1}^{k} u_{i,j} = 1, \text{ and } u_{i,j} \in [0, 1], 1 \leq i \leq n, \tag{2}$$

where $U = [u_{i,j}]$ is an $n \times k$ matrix, $u_{i,j}$ represents the degree of membership of ith object x_i to the jth cluster z_j, $Z = [z_1, z_2, \ldots, z_k]^T$ is an $k \times m$ matrix containing the cluster centers, and $D_{i,j}$ is a dissimilarity measure between the jth cluster center and the ith object. Here, the square of the eucledian norm is used as the dissimilarity measure, i.e.

$$D_{i,j} = \sum_{l=1}^{m} \left(Z_{j,l} - X_{i,l}\right)^2 \tag{3}$$

and λ represents the weighting exponent $(\lambda > 1)$.

The usual method toward optimization of P in (1) is to use the partial optimization for U and Z. In this method, we first fix U and minimize the reduced P with respect to Z. Then, we fix Z and minimize the reduced P with respect to U.

Given U fixed, Z is updated as

$$Z_{i,j} = \frac{\sum_{i=1}^{n} \left(u_{i,j}\right)^{\lambda} x_{i,l}}{\sum_{i=1}^{n} \left(u_{i,j}\right)^{\lambda}} \tag{4}$$

Given that Z fixed, U is updated as follows:

$$u_{i,j} = \frac{D_{i,j}^{(\lambda-1)}}{\sum_{i=1}^{k} \left(D_{i,j}^{(\lambda-1)}\right)} \tag{5}$$

The Fuzzy *k*-Means clustering algorithm is summarized as follows.

2.1.1 Fuzzy k-Means Algorithm:

1. Randomly choose initial points $U^{(0)}$. Determine $Z^{(0)}$ such that $P(U^{(0)}, Z^{(0)})$ is minimized by using (6).set $t = 0$.
2. Let $\hat{Z} = Z^t$, update U^t to U^{t+1} by using (7). If $P\left(U^{t+1}, \hat{Z}\right) = P\left(U^t, \hat{Z}\right)$, output $\left(U^t, \hat{Z}\right)$ and stop; otherwise, go to step 3.
3. Let $\hat{U} = U^{t+1}$, update Z^t to Z^{t+1} by using (6). If $\left(\hat{U} = Z^{t+1}\right) = P\left(\hat{U}, Z^t\right)$, output $\left(\hat{U}, Z^t\right)$ and stop; otherwise, set $t = t+1$ and go to step 2.

The clustering information obtained from the fuzzy *k* means clustering is used to find the membership degree for each input LR patch.

2.2 *Fuzzy Clustering Based Neighbor Embedding for SR Reconstruction*

The neighbor embedding algorithm was first used for super resolution by Chang et al. in SRNE [6]. They use this algorithm for estimating HR patches from corresponding LR patches. Local Linear Embedding (LLE) is based on the idea that local geometry of each HR patch can be characterized by the reconstruction weight with the data points reconstructed from its neighbors.

The fuzzy neighbor embedding algorithm [6, 18, 19] for the proposed method can be summarized as follows.

2.2.1 Fuzzy Clustering Based Neighbor Embedding Algorithm:

1. For each patch x_t^q in input image X_t:

a. Find the cluster membership degree for that patch using the cluster model constructed from the training LR patches. The membership degree is computed as follows.

$$u_{q,j} = \frac{D_{i,j}^{(\lambda-1)}}{\sum_{i=1}^{k} D_{i,j}^{(\lambda-1)}} \tag{6}$$

where λ is weighting exponent and $D_{i,j}$ is a dissimilarity measure between the jth cluster center and the qth input LR patch and is computed as follows.

$$D_{q,j} = \sum_{l=1}^{m} \left(Z_{j,l} - Y_{q,l}\right)^2 \tag{7}$$

b. Find the eucledian distance between x_t^q and all the training LR patches x_s^q. The distance matrix $Dist_{i,j}$ is computed as follows.

$$Dist_{i,j} = \sum_{i=1}^{m} (x_s(i,l) - x_t(i,l))^2 \tag{8}$$

c. The distance matrix $Dist_{i,j}$ between x_s^i and x_t^j is tuned using the membership degree associated with x_s^i as follows.

$$Dist\prime_{i,j} = Dist_{i,j} + \max(Dist)U_{i,k} \tag{9}$$

where $U_{i,k}$ is the membership degree associated with the x_s^i for the cluster k. The cluster k is the cluster in which x_t^j has largest membership degree.

d. Find the set $N_q$of K nearest neighbors in X_s
e. Compute the reconstruction weights of the neighbors that minimize the error of reconstructing x_t^q.The weight is computed as follows.

$$w_q = \frac{G_q^{-1}\,1}{1^T\,G_q^{-1}\,1} \tag{10}$$

where G_q is the local gram matrix for x_t^q and it is defined as

$$G_q = \left(x_t^q\,1^T - X\right)^T \left(x_t^q\,1^T - X\right) \tag{11}$$

where 1 is column vector of ones and X is a m × K matrix with its column being the neighbors of x_t^q in X_s.

f. The high resolution embedding y_t^q of x_t^q is constructed as the weighted sum of high resolution features of *K* nearest neighbor in X_s.

$$y_t^q = \sum_{x_s^p} w_q\, y_s^p \tag{12}$$

2. Construct the target high resolution image Y_t by combining the high resolution patches obtained from the step 1. We inforce the local compatibility and smoothness constraints between the adjacent patches by averaging feature values in the overlapped region.

When an input LR image is given, first the input image is transformed into YIQ color space. The luminance value extracted from the YIQ model is used for HR image reconstruction using neighbor embedding while the other two components are just interpolated to the desired resolution. Then the image is decomposed into patches and each patch is represented by its feature vector. Then the HR image is reconstructed using Fuzzy clustering based neighbor embedding described in Sect. 2.

3 Experimental Results and Discussion

The performance of the proposed method is evaluated quantitatively and qualitatively and the comparison with Bicubic, SRNE, CSNE is given in this section.

Fig. 1 Training images

Fig. 2 Input LR images. From *left* to *right*, *top* to *bottom* are labeled No. 1 to No. 6

Table 1 Comparison of different algorithms using PSNR(dB)

Testing image	Methods			
	Bicubic	SRNE	CSNE	Proposed
No. 1	30.0710	30.1815	30.2760	30.4950
No. 2	31.2954	31.5645	31.5681	31.6194
No. 3	29.5992	29.6635	29.6919	29.7983
No. 4	29.9147	29.8887	29.9539	29.9704
No. 5	33.2656	33.6616	33.7431	33.8435
No. 6	33.5677	34.0272	34.0611	34.0912

In example based super resolution, the selection of training set is very important for the reconstruction quality of HR image. For testing the performance of the proposed method quantitatively, the LR images (both for training and testing) are constructed form HR images by degrade the HR image using an average filter within a 4×4 neighborhood and down sampled by a factor 4. Figure 1 shows the training images used in the experiment. Figure 2 shows the test LR images used in our experiment labeled from No1 to No 8. As that of CSNE [12], the low-resolution patch size is 3×3, and there are 2 pixels overlapped between the adjacent patches. For scaling with a factor 4, high resolution patch size is set to 12×12 and there are 8 pixels overlapped between the adjacent patches. The proposed algorithm uses the same features to represent the low resolution and high resolution patches as CSNE does.

Peak-signal-to-noise ratio (PSNR) is used as the metric to evaluate the super resolution results. The PSNR is defined as

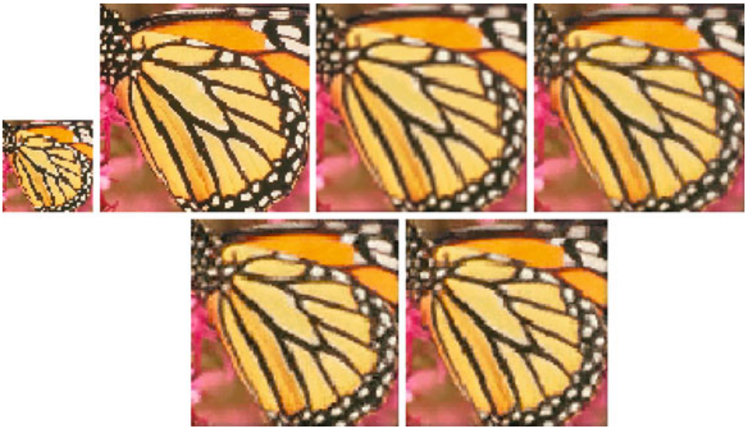


Fig. 3 4× recovery of No. 1 using different methods. From *left* to *right*, *top* to *bottom*: the low resolution image; the original image (ground truth); Bicubic(30.0710 dB); SRNE(30.1815 dB); CSNE(30.2760 dB); the proposed method(30.4950 dB)

Fig. 4 4× 4× recovery of No. 3 using different methods. From *left* to *right*, *top* to *bottom*: the low resolution image; the original image (ground truth); Bicubic(29.5992 dB); SRNE(29.6635 dB); CSNE(29.6919 dB); the proposed method(29.7983 dB)

$$\mathrm{PSNR} = 10\log_{10}\frac{255^2}{\mathrm{MSE}} \tag{13}$$

where MSE is the mean squared error and is defined as follows.

$$\mathrm{MSE} = \frac{1}{\mathrm{n}}\sum_{\mathrm{i}=1}^{\mathrm{n}}(\widehat{y}_{\mathrm{i}} - y_{\mathrm{i}})^2 \tag{14}$$

where $\widehat{y}_i$ stands for the value of pixel in the original image Y and y_i stands for the values of corresponding pixels in output Y_t. And n represents the total number of pixels in the image. Table 1 shows the PSNR of different methods for the test images, including the proposed method. The proposed method gives good results for all the test images than the other algorithms.

The 4× recovery of LR version of No.1 and No.3are shown in Figs. 3 and 4 respectively. Qualitatively, it is clear that the proposed method is able to recover the better visual quality than the other algorithms.

4 Conclusion

Using the idea of CSNE [12], a novel example based super resolution technique is presented in this paper. In this method the concept of fuzzy clustering is used in neighbor embedding. The class information is learnt from the LR training patches using Fuzzy K- Means Clustering. And this class information is further used to find the k nearest LR patches by adjusts the distance between the input LR patch and training LR patches. Then using the neighbor embedding algorithm, target HR patches are constructed as the optimal weighted sum of training HR patches corresponds to the LR training patches. The experimental results show that the proposed algorithm achieves better results than other neighbor embedding algorithms.

References

1. Park S, Park M, Kang MG (2003) Super-resolution image reconstruction, a technical overview. IEEE Signal Process Mag 20(5):21–36
2. Huang TS, Tsai RY (1984) Multi-frame image restoration and registration. Adv Comput Vis Image Process 1:317–339
3. Elad M, Feuer A (1997) Restoration of a single super resolution image from several blurred, noisy, and under sampled measured images. IEEE Trans Image Process 6(12):1646–1658
4. Freeman WT, Pasztor EC, Carmichael OT (2000) Learning low-level vision. Int J Comput Vis 40(1):25–47
5. Freeman WT, Jones TR, Pasztor EC (2002) Example-based super-resolution. IEEE Comput Graph Appl 22(2):56–65
6. Chang H, Yeung DY, Xiong Y (2004) Super-resolution through neighbor embedding. In: Proceedings of IEEE Computer Society Conference on Computer Vision Pattern Recognition, pp 275–282
7. Geng B, Taoet D et al (2009) Ensemble manifold regularization. In: Proceedings of IEEE Computer Society Conference Computer Vision Pattern Recognition Workshops, CVPR Workshops, pp 2396–2402
8. Roweis ST, Saul LK (2000) Nonlinear dimensionality reduction by locally linear embedding. Science 290(5500):2323–2326
9. Li X, Lin S, Yan S, Xu D (2008) Discriminant locally linear embedding with high-order tensor data. IEEE Trans Syst Man Cybern B 38(2):342–352

10. Zhou T, Tao D, Wu X (2010) Manifold elastic net: a unified frame-work for sparse dimension reduction. Data Mining Knowl Discov 22(3):340–371
11. Chan TM, Zhang J, Pu J, Huang H (2009) Neighbor embedding based super-resolution algorithm through edge detection and feature selection. Pattern Recogn Lett 30(5):494–502
12. Zhang K, Tao D et al (2011) Partially supervised neighbor embedding for example-based image super-resolution. IEEE J Selected Topics Sig Process 5(2)
13. Bouman CA (2005) Cluster: an unsupervised algorithm for modeling Gaussian mixtures. Purdue Univ West Lafayette, IN (On-line)
14. Zeng H, Cheung Y (2009) A new feature selection method for Gaussian mixture clustering. Pattern Recogn 42(2):243–250
15. Hoppner F, Klawonn F, Kruse R (1999) Fuzzy cluster analysis: methods for classification. Data analysis and image recognition. Wiley, New York
16. Krishnapuram R, Frigui H, Nasraoui O (1995) Fuzzy and possibilistic shell clustering algorithms and their application to boundary detection and surface approximation—part I and II. IEEE Trans Fuzzy Syst 3(4):29–60
17. Hoppner F (1997) Fuzzy shell clustering algorithms in image processing: fuzzyc-rectangular and 2-rectangular shells. IEEE Trans Fuzzy Syst 5(4):599–613
18. Ruspini ER (1969) A new approach to clustering. Info Control 19:22–32
19. Bezdek JC (1980) A convergence theorem for the fuzzy ISODATA clustering algorithms. IEEE Trans Pattern Anal Mach Intell 2:1–8

Fingerprint Authentication System Based on Minutiae and Direction Field Toning Technique

S. Valarmathy, M. Arunkumar and M. Padma

Abstract Fingerprint authentication refers to the automated method of verifying a match between two human fingerprints. Fingerprints are one of many forms of biometrics used to identify the individuals and verify their identity. Minutiae points are local ridge characteristics which can provide unique information. The conventional methods have utilized this minutiae information only as a point set. As a global feature, orientation field has extremely higher inter-personal variation than that of minutiae points. This proposed method describes the implementation of an Automatic Fingerprint Identification System (AFIS) which incorporates both local and global features of the fingerprints and operates in three stages: (1) minutiae extraction (2) reconstruction of orientation field and (3) fusion matching. An Improved Gabor filter algorithm is used for reliable minutiae extraction. From this extracted minutiae, the orientation field is reconstructed by using Interpolation-Model based (IM) method and utilized in the matching stage. This reconstructed orientation field matching is then fused with the minutiae-based matching. Thus the proposed scheme can enhance the performance of the system and obtain better matching accuracy than conventional methods.

Keywords Decision-level fusion · Interpolation · Minutiae · Orientation field · Polynomial model

S. Valarmathy (✉) · M. Arunkumar · M. Padma
Bannari Amman Institute of Technology, Erode, India
e-mail: atrmathy@gmail.com

M. Arunkumar
e-mail: mak5phd@gmail.com

M. Padma
e-mail: mpadmaece@gmail.com

Mohan S. and S. S. Kumar (eds.), *Proceedings of the Fourth International Conference on Signal and Image Processing 2012 (ICSIP 2012)*, Lecture Notes in Electrical Engineering 222,
DOI: 10.1007/978-81-322-1000-9_7,

1 Introduction

The term Biometrics relates to the measurement (metric) of characteristics of a living (Bio) thing in order to identify a person. Among many Biometric recognition systems, Fingerprint-based identification has been successfully used in forensic and civilian applications such as criminal identification, access control and ATM card verification, due to the uniqueness of fingerprints.

The fingerprint is formed in third and fourth month of fetal development. A fingerprint is made up of a series of ridges and furrows on the surface of the finger. The minutiae are ridge endings or bifurcations on the fingerprints. Most fingerprint recognition systems store the minutiae template in the database.

This kind of minutiae-based recognition systems consists of two steps, i.e., minutiae extraction and minutiae matching.

The minutiae feature of a given fingerprint is compared with the minutiae template, and the matched minutiae will be found out. If the matching score exceeds a predefined threshold, the two fingerprints can be regarded as belonging to a same finger. However, these kind of methods cannot provide enough distinguishing abilities for large-scale fingerprint identification tasks. Obviously, a better usage of minutiae is very important for fingerprint recognition systems.

2 Related Work

J. Gu, J. Zhou and C. Yang [1] have proposed a novel representation for fingerprints which has included both minutiae and model-based orientation field. J. Gu, J. Zhou and D. Zhang [2] have established a novel model for the orientation field of fingerprints which has improved the performance of orientation estimation. A. Jain and L. Hong [3] have described the design and implementation of an on-line fingerprint verification system. For minutiae matching an alignment-based elastic matching algorithm has been developed.

A. Jain, S. Prabhakar, and L. Hong [4] have presented a finger code algorithm that has reduced the fingerprint matching time for a large database. But this algorithm suffers from the requirement that the region of interest be correctly located.

S. Pankanti, S. Prabhakar and A. K. Jain [5] have addressed the problem of fingerprint individuality by quantifying the amount of information available in minutiae features to establish a correspondence between two fingerprint images.

J. Qi, S. Yang and Y. Wang [6] have defined a novel feature vector for each fingerprint minutia based on the global orientation field.

A. Ross, J. Shah, A. K. Jain [7] have showed that three levels of information about the parent fingerprint could be elicited from the minutiae template alone, viz., (1) the orientation field information (2) the class or type information and (3) the friction ridge structure. T.P.Weldon, W. E. Higgins and D. F. Dunn [8] have proposed a Gabor filter based method for fingerprint recognition. J. Zhou and J. Gu [9] have proposed a model-based method for the computation of orientation field.

2.1 Proposed Method

In this paper, an improved version of the algorithm called Gabor filter algorithm is proposed for minutiae extraction. This algorithm can adaptively improve the clarity of the ridge and valley structures of the input fingerprint images. For reconstruction of orientation field, two different methods are available.

They are the following:

- Interpolation method
- Polynomial model-based method

Interpolation algorithm is used to estimate the orientation field from minutiae template, in which the orientation of a given point is computed from its neighboring minutiae. Polynomial model based method is used to map the orientation field to a continuous complex function.

In this paper, a method called Interpolation-Model based (IM) method is proposed which is the combination of Interpolation and Polynomial model- based method. Firstly, we interpolate a few "virtual" minutiae in the sparse areas, and then apply the model- based method on these mixed minutiae (including the "real" and "virtual" minutiae). After that, the reconstructed orientation field is used into the matching stage by combining with conventional minutiae-based matching. Figure 1 shows then flowchart of the proposed method.

The paper is organized into the following sections. Section 2 is an overview of related work. The minutiae extraction process is described in Sect. 3. Section 4 discusses the reconstruction of orientation field. Decision level fusion matching is proposed in Sect. 5 and results and conclusion are contained in Sects. 6 and 7 respectively.

3 Minutiae Extraction

Extraction of true minutiae consists of a series of processing stages. A gray level fingerprint image is considered to test the improved Gabor filter algorithm (Fig. 2).

The first stage is the De-noising of the input fingerprint image. Noise elimination is defined as the process of removing undesired pixels in the image. The second stage of processing is the Histogram equalization. Histogram equalization is defined as the process of obtaining an image with gray levels having uniform density.

The third and important stage of processing is the improved Gabor filter enhancement. This enhancement process is used to improve the image quality by removing noise using band pass filter. The next stage of processing is binarization. Binarization is defined as the process of converting an enhanced gray level image into a binary image. The next stage is the thinning process. Thinning is a morphological operation that successively erodes away the foreground pixels until

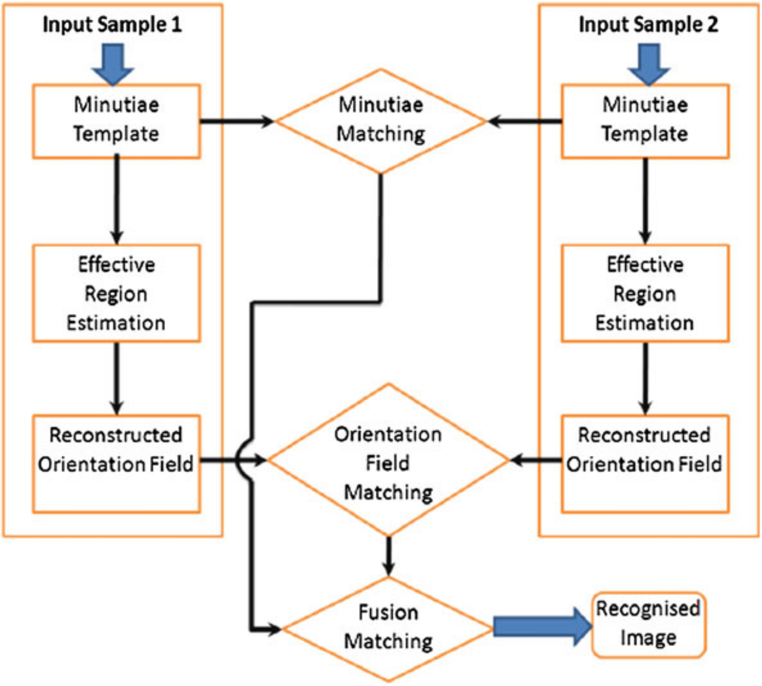


Fig. 1 Flowchart of the proposed scheme

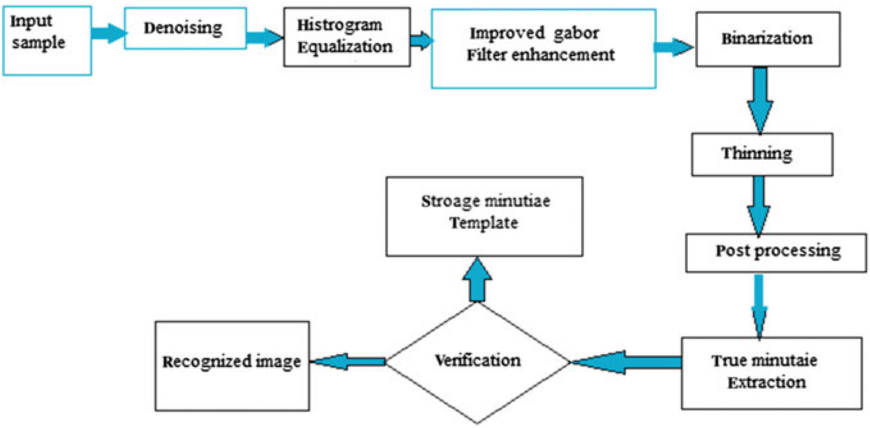


Fig. 2 Flowchart of the minutiae extraction process

they are one pixel wide. The next stage is the post processing stage. Post processing is defined as the process of removing redundant minutiae. After completions of post processing, true minutiae are extracted The extracted minutiae are compared with the minutiae templates which have been stored. If the matching score exceeds a predefined threshold, then the two fingerprint images can be regarded as belonging to the same finger.

4 Reconstruction of Orientation Field

4.1 Estimating the Effective Regions

The effective region can extract by only using minutiae information but the ridges and valleys. In this situation, we can extract the effective region by finding the smallest envelope that contains all the minutiae points (Fig. 3).

4.2 Interpolation

The minutiae of a fingerprint always distribute "non-uniformly," which results in some sparse regions where there are few minutiae. In order to give high weights to the minutiae in the sparse area, we produce "virtual" minutiae by using interpolation before modeling.

Since the orientation field of fingerprints always change smoothly, it is possible to estimate the direction of a point by examining the direction of minutiae points in the local region. In order to interpolate "virtual" minutiae in the sparse area, we

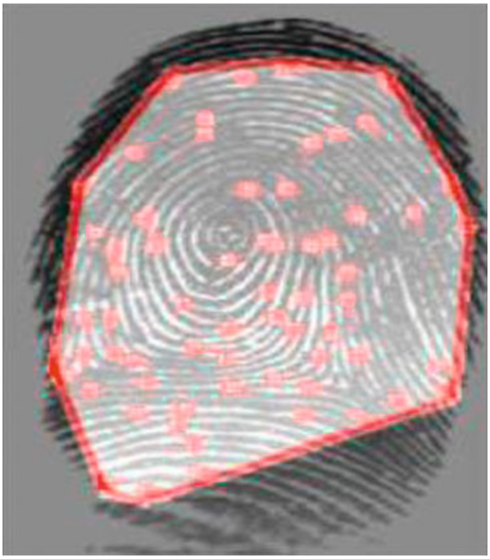


Fig. 3 Illustration of effective region estimation

choose three minutiae points to construct a triangle, and estimate the orientation field in the triangle by these three minutiae. The algorithm has the following two main steps.

4.2.1 Triangulation

The fingerprint is divided into many triangles. Consider a set of N ($N \geq 3$) points in the plane, the simplest way to triangulate them is to add to the diagonals from the first point to all of the others. However, this has the tendency to create skinny triangles. Skinny triangles have to be avoided, or equivalently, small angles in the triangulation. Delaunay triangulation has been used which minimizes the maximum angle over all possible triangulations and it can be constructed in n log n time. The triangles have no intersection, so each point can be covered by only one triangle, as shown in Fig. 5b.

4.2.2 Producing "Virtual" Minutiae Using Interpolation

Let P (x, y) denotes the "virtual" minutiae located inside the triangle, Δ M1 M2 M3, $d_i = \| P - M_i \|$ be the Euclidean distance of these "virtual" minutiae from the *i*th vertex M_i. And let θ_i be the direction corresponds to the vertex, M_i. It is clear that a vertex should affect more on the "virtual" minutiae P if it is closer to P than other vertex.

Thus, the direction of the pixel P, θ_P is estimated as in (1–4)

$$\theta_i' = \begin{cases} \theta_i, & 0 \leq \theta_i < \pi \\ 2\pi - \theta_i, & 0 \leq \theta_i < 2\pi \end{cases} \tag{1}$$

Since θ_1, θ_2 and θ_3 are rotationally symmetric, we can (3) assume that $\theta_1' \leq \theta_2' \leq \theta_3'$. For example, if' $\theta_1' \geq \theta_2' \geq \theta_3'$, we can exchange the denotation: change θ_1' to θ_2', θ_2' to θ_3' and θ_3' to θ_1'.

$$\begin{cases} \theta_1'' = \begin{cases} \theta_1' + \pi \; if \; \theta_2' > \dfrac{\pi}{2} \; and \; \theta_3' - \theta_1' > \dfrac{\pi}{2} \\ \theta_1'', \; otherwise \\ \theta_2'' = \theta_2' \end{cases} \\ \theta_3'' = \begin{cases} \theta_3' - \pi \; if \; \theta_2' < \dfrac{\pi}{2} \; and \; \theta_3' - \theta_1' < \dfrac{\pi}{2} \\ \theta_3', \; otherwise \end{cases} \end{cases} \tag{2}$$

The ridge line orientation is defined in the range of $[0, \pi]$. (2) takes into account that π phase jumps may appear in the estimation.

$$\theta p' = \frac{d_2 d_3}{d_1 d_2 + d_2 d_3 + d_3 d_1}\theta_1'' + \frac{d_3 d_1}{d_1 d_2 + d_2 d_3 + d_3 d_1}\theta_2'' + \frac{d_1 d_2}{d_1 d_2 + d_2 d_3 + d_3 d_1}\theta_3'' \tag{3}$$

$\theta_p{}''$ is calculated as,

$$\theta_p'' = \begin{cases} \pi + \theta_p' & if\ -\frac{\pi}{2} < \theta_p' < 0 \\ \theta_p' & if\ 0 \le \theta_p' < \pi \\ 2\pi - \theta_p' & if\ 0 \le \theta_p' < \frac{3\pi}{2} \end{cases} \tag{4}$$

The computation is illustrated in Fig. 4. After the interpolation step, the minutiae distribute "uniformly" as shown in Fig. 5c.

4.3 Polynomial Model-Based Method

The orientation field is firstly mapped to a continuous complex function. Denoting θ(x, y) and U(x, y) as the orientation field and the transformed function, respectively, the mapping can be defined as

$$\begin{aligned} \mathrm{U(x,\ y)} &= \mathrm{RE\ (x,\ y) + i.\ IM\ (x,\ y)} \\ &= \cos 2\theta\,\mathrm{(x,\ y)} + \mathrm{i.}\ \sin 2\theta\,\mathrm{(x,\ y)} \end{aligned} \tag{5}$$

where RE(x, y) and IM(x, y) denote respectively the real part and imaginary part of the complex function, U(x, y).

Obviously, RE(x, y) and IM(x, y) are continuous with x, y, in those regions. The above mapping is a one-to-one transformation and θ(x, y) can be easily reconstructed from the values of RE(x, y) and IM(x, y). To globally represent these, two

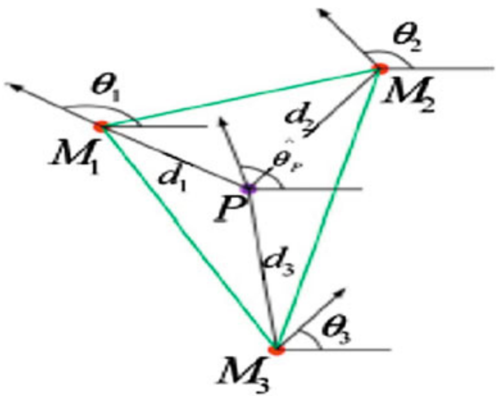


Fig. 4 Computation of a pixel in a triangle

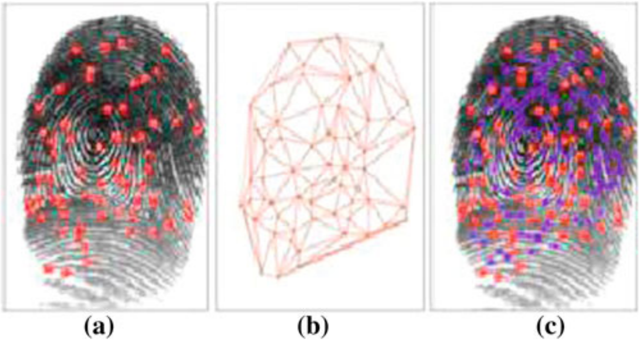


Fig. 5 Interpolation step: (**a**) The Minutiae image; (**b**) The triangulated image; (**c**) Virtual minutiae by interpolation (the bigger red minutiae are "real," while the smaller purple ones are "virtual")

bi variate polynomial models are established, which are denoted by PR(x, y) and PI(x, y), respectively. These two polynomials can be formulated as

$$\mathrm{PR}(x,\ y) = X^T \cdot p_1 \cdot \mathrm{Y} \tag{6}$$

$$\mathrm{PI}(x,\ y) = X^T \cdot p_1 \cdot \mathrm{Y} \tag{7}$$

where $X = (1,\ x,\ x^2 \ldots\ldots x^N)^T$ and $Y = (1,\ y,\ y^2 \ldots\ldots y^N)^T$. In these two formulas, N is the order of the polynomial model. There are $(N + 1) \times (N + 1)$ parameters of p_1 and p_2 which need to be calculated. Computing the parameters is a fitting process. Using square sum error for evaluation, the formula becomes

$$\left(p_1^*, p_2^*\right) = \arg\min_{p_1 p_2} \sum\nolimits_{(x,y)\in\Omega} \left[(PR(x,y,p_1) - \cos 2\theta_0(x,y))^2 + (PI(x,y,p_2) - \sin 2\theta(x,y))^2\right] \tag{8}$$

where Ω is the set of the effective region and θ_0(X,Y) the original orientation field.

5 Decision-Level Fusion Matching

To compare two fingerprints orientation field, the first step is alignment of these two fingerprints. The Hough transform based approach has been chosen to finish the alignment due to its simplicity. In the matching step, the correlation between two aligned orientation fields, A and B is computed as below. Let Ω denotes the intersection of the two effective regions after alignment, and N is the total number of points in Ω. The matching score between two orientation fields is defined as

$$s(A, B) = \frac{1}{N} \sum\nolimits_{(i,j\in\Omega)} \delta(i,j) \tag{9}$$

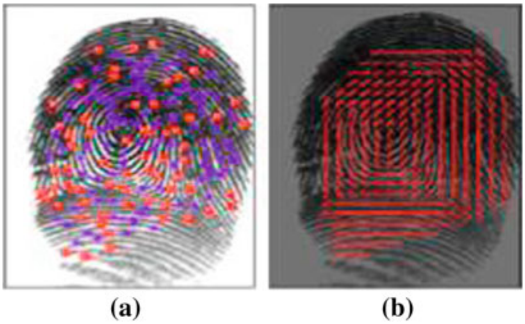


Fig. 6 Results of the proposed algorithm: (**a**) Virtual minutiae by interpolation (the bigger red minutiae are "real", while the smaller purple ones are "virtual"); (**b**) The reconstructed orientation field

In (9), $\delta(i,j)$ is the difference between the orientation values at the point (i, j) in image A and B, which is formulated as follows:

$$\partial(i,j) = \begin{cases} \delta_0(i,j), & \text{if } \delta_0(i,j) \leq \dfrac{\pi}{2} \\ \pi - \delta_0(i,j), & \text{otherwise} \end{cases}$$

$$\delta_0(i,j) = \theta_A(i,j) - \theta_B(i,j)$$

where θ_A(i,j) and θ_B(i,j)are the direction of point (i, j), in image A and B. If the matching score s (A, B) is higher than a certain threshold, we say the two orientation fields are "matched." (Fig. 6).

The result shows that the model-based method works better than the Interpolation method and the interpolation-and-model algorithm performs the best.

As an example, in Fig. 7a, there is a minutiae M (marked with ellipse) whose direction is estimated wrongly, and Fig. 7b shows the corresponding poor result (marked with ellipse) by interpolation. IM method can consider all of the minutiae information, so the effect of the wrong direction of M can be reduced by other

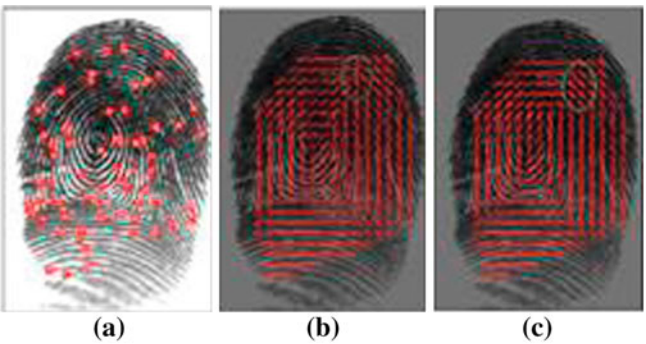


Fig. 7 Comparison result I: (**a**) Minutiae image with a wrong direction (marked with ellipse); (**b**) The corresponding poor result by interpolation (marked with ellipse); (**c**) The corresponding good result by the proposed IM method

correct minutiae. Thus, using the proposed method to reconstruct orientation field can overcome the problem induced by M to some extend by adding some positive contribution of other minutiae. Figure 7c shows the orientation field reconstructed based on IM method which can overcome the problem induced by the wrongly estimated minutiae M (marked with ellipse).

6 Conclusion

Orientation field is important for fingerprint representation. In order to utilize the orientation information in automatic fingerprint recognition systems which only stores minutiae feature, a novel method has been proposed to utilize the minutiae for fingerprint recognition. The reconstructed orientation field information has been utilized into the matching stage. The proposed algorithm combines the interpolation method and model-based method to reconstruct orientation field, and reduces the effect of wrongly detected minutiae. Hence the performance of the Proposed method is better than existing methods for real time applications. Our future work is currently going on calculating FAR, FRR and EER for different databases.

References

1. Gu J, Zhou CJ, Yang C (2006) Fingerprint recognition by combining global structure and local cues. IEEE Trans Image Process 15:1952–1964
2. Gu J, Zhou J, Zhang D (2004) A combination model for orientation field of fingerprints. Pattern Recog 37:543–553
3. Jain AK, Hong L, Bolle R (1997) On-line fingerprint verification. IEEE Trans Pattern Anal Mach Intell 19:302–314
4. Jain AK, Prabhakar S, Hong L (1999) A multichannel approach to fingerprint classification. IEEE Trans Pattern Anal Mach Intell 21:348–359
5. Pankanti S, Prabhakar S, Jain AK (2002) On the individuality of fingerprints. IEEE Trans Pattern Anal Mach Intell 24:1010–1025
6. Qi J, Yang S, Wang Y (2005) Fingerprint matching combining the global orientation field with minutia. Pattern Recogn Lett 26(15):2424–2430
7. Ross A, Shah J, Jain AK (2007) From template to image: reconstructing fingerprints from minutia points. IEEE Trans Pattern Anal Mach Intell 29:544–560
8. Weldon TP, Higgins WE, Dunn DF (1996) Efficient Gabor filter design for texture segmentation. Pattern Recogn 29(12):2005–2015
9. Zhou J, Gu J (2004) A model-based method for the computation of fingerprints orientation field. IEEE Trans Image Process 13:821–835
10. Maio D, Maltoni D, Cappelli R, Wayman JL, Jain AK (2002) FVC2002: second fingerprint verification competition. In: Proceeding international conference on pattern recognition, August 2002, vol 3, pp 811–814

Segment-Based Stereo Correspondence of Face Images Using Wavelets

C. J. Prabhakar and K. Jyothi

Abstract In this paper, we introduce color segmentation based stereo correspondence for face images using wavelets. The intensity based correlation techniques are commonly employed to estimate the similarities between the stereo image pair, sensitive to shift variations and relatively lower performance in the featureless regions. Therefore, instead of pixel intensity, we consider wavelet coefficients of an approximation band, which is less sensitive to the shift variation. The approximation subband of reference image is segmented using mean shift segmentation method. A self-adapting dissimilarity measure that combines sum of absolute differences of wavelet coefficients and a gradient is employed to generate a disparity map of the stereo pairs. In our method instead of assigning a disparity value to a pixel, a disparity plane is assigned to each segment. Results show that the proposed technique produces smoother disparity maps with less computation cost.

Keywords Stereo matching · Discrete wavelet transform · Disparity plane · Segmentation

C. J. Prabhakar (✉)
Department of Studies in Computer Science, Kuvempu University, Shankaraghatta, Karnataka, India
e-mail: psajjan@yahoo.com

K. Jyothi
Department of IS&E, J.N.N College of Engineering, Shimoga, Karnataka, India
e-mail: jyothi6575@rediffmail.com

Mohan S. and S. S. Kumar (eds.), *Proceedings of the Fourth International Conference on Signal and Image Processing 2012 (ICSIP 2012)*, Lecture Notes in Electrical Engineering 222,
DOI: 10.1007/978-81-322-1000-9_8,

1 Introduction

The stereo correspondence is process of finding the corresponding points in the stereo image pair. The estimated disparity between the stereo images is used to extract depth information of the scene. The depth information of the scene is the specific step for 3D object reconstruction, 3D face reconstruction and industrial automation systems. In computer vision, estimation of the stereo correspondence between the stereo image pair is one of the critical problems. The stereo correspondence is mainly categorized into two types; those are local search method [1–7] and global search method [8, 9]. Local search stereo correspondence uses the information within the small area, while global search methods optimize some of the global (energy) function.

Some of the local search methods, such as block matching, gradient based optimization and feature matching are very efficient. The computation time of the local-based (area) methods are very less, but produces less accurate results. While global-based (energy) method consumes more time for computation and generates a more accurate results. The most commonly used area based matching algorithms are the sum of absolute differences (*SAD*), the sum of squared differences (*SSD)* and the normalized *SSD.*

Muhlman et al. [1] introduced a local search based matching technique for RGB stereo images. This method uses left-to-right consistency and uniqueness constrain to produce the disparity map. The resulting disparity map is smoothed by applying a median filter. Yoon et al. [2] proposed a correlation-based local stereo correspondence matching technique, which uses a refined implementation of the Sum of Absolute Differences (SAD) criteria and a left-to-right consistency check to minimize the mismatch errors. This algorithm uses a variable correlation window size to reduce the errors in the areas containing blurring or mismatch errors. Yoon and Kweon [3] have presented another area-based algorithm, which uses adoptive weights based on the color similarity and geometric distances for each pixel in the search area to reduce the ambiguity errors. Yong and Lee [4] employed area-based method, which is insensitive to radiometric variation between stereo images using adaptive normalized cross correlation method. Hamzah et al. [5] have proposed block matching algorithm to stereo correspondence using sum of absolute difference of the intensity. In multiple-window methods [6, 7] select an optimal support window from predefined multiple windows, which are placed at different positions with the same shape. Fusiello et al. [6] performed the correlation with nine different windows for each pixel and consider the disparity with the smallest matching cost. Kang et al. [7] also presented a multiple-window method that examines all windows containing the pixel of interest.

All the methods mentioned above have a common limitation. To find the optimal support window with a different shape and size is very difficult. For this reason, the methods limit their search space by constraining the shape of a support window. Rectangular and constrained-shaped windows may be inappropriate for pixels near arbitrarily shaped depth discontinuities. This problem can solve by segmentation-based methods [10–14], which uses segmented regions with arbitrary sizes and shapes as support windows. In this approach, it is based on the assumption that the scene structure can be approximated by a set of non-overlapping planes in the disparity space and that each plane is coincident with at least one homogeneous color segment in the reference image.

Most of the stereo matching methods working in the spatial domain, assume that the projection of an object will have the same area in both images. However, this condition is violated by perspective projection and causes the ambiguous mismatches or false targets. This will greatly reduce the accuracy of a stereo vision system and can be avoided by considering the stereo matching in a frequency domain.

Current research in stereo correspondence estimation has attracted a lot of focus on multiresolution techniques, based on wavelets/multiwavelets scale-space representation and analysis [17]. However, very little work has been reported in this regard. Wavelet based stereo matching algorithms have received much attention due to scale-space localization properties of the wavelets [15–18]. In this paper, a wavelet-based stereo matching algorithm using a self-adapting dissimilarity measure technique is presented. A discrete wavelet transform is applied to the rectified stereo images to decompose them into subbands. Approximation subband of reference image is segmented using mean shift segmentation method [19]. The self-adapting dissimilarity measure is used to generate a disparity map.

The remaining sections of the paper are organized as follows. In Sect. 2, a brief description of wavelet transforms is presented. The proposed wavelet based stereo matching method is discussed in Sect. 2. The experimental results are presented in Sect. 3. Finally, the conclusion is given in Sect. 4.

2 Stereo Matching

In the proposed stereo matching technique, stereo image pair is input to the system. The images are first rectified using uncalibrated rectification method to suppress the vertical displacement [20]. A 2-D discrete wavelet transform (DWT) is used to decompose the given image into four subbands, namely: (1) LL; (2) LH; (3) HL and (4) HH. Approximation subband of the input image has different spectral content, while the detail subbands mainly contain a mixture of horizontal, vertical and diagonal details of input image. In addition to this, the information in the base bands is less sensitive to the shift variability of the wavelets. Approximation subband of reference image is segmented using mean shift segmentation method. The self-adapting dissimilarity measure [21] is used for stereo matching.

2.1 Wavelets Fundamentals

Wavelet transforms are based on small waves, called wavelets of varying frequency and limited duration. In discrete wavelet transform (DWT), an image is decomposed into a set of band limited components, called subbands. Wavelet theory is based on the refinement equations are given below

$$\phi(t) \ = \ \sum_h w_h \, \phi\,(Mt - h), \tag{1}$$

$$\varphi(t) = \sum_h c_h \, \phi\,(Mt - h). \tag{2}$$

where c_h and w_h represents the scaling and wavelet coefficients. ϕ(t) is wavelet function and φ(t) is scaling function. Discrete wavelet transform is consisting of a scaling function and three wavelet functions. *M represents* the subband number. The scaling function and wavelet function has finite support, if and only if the most of coefficients w_h and c_h are finite. Multiwavelet consists of several wavelet and scaling functions and are defined as:

$$\phi(t) = [\phi_1(t), \phi_2(t), \ldots, \phi_k(t)]^T, \tag{3}$$

$$\varphi(t) = [\varphi_1(t), \varphi_2(t), \ldots, \varphi_k(t)]^T, \tag{4}$$

where $\phi(t)$ and $\varphi(t)$ are the multiwavelet functions and multiscaling, with k scaling and k wavelet functions. Wavelet transforms results in the information belonging to the approximation space A_k and detail space V_k possessing the property

$$V_{k+1} = A_k \oplus V_k. \tag{5}$$

Mallat [15] introduced one of the most widely used discrete wavelet transform. In Mallat's representation, the details space (V_k) consists of three components, which are horizontal, vertical and diagonal.

2.2 Color Segmentation

The first step of estimation of the disparity map is to divide the approximation subband of reference image into regions of homogeneous color using mean shift segmentation. We assume that disparity variation at that region is negligible. The edge information is also incorporated in this segmentation method. The mean shift color segmentation is essentially defined as a gradient ascent search for maxima in a density function defined over a high dimensional feature space.

2.3 Estimation of Disparity

In our proposed method, the disparity is estimated by self adaptive dissimilarity measures, that conjugate sum of absolute differences of wavelet coefficients of stereo pairs and a relative changes in the color of images. That is defined as

$$SAD\,(x, y, d) \;=\; \sum_{(i,j)\,\in N(x,y)} f_1(i,j) - f_2(i + d,j). \tag{6}$$

$$\begin{aligned} Gradient\,(x, y, d) \;=\; & \sum_{(i,j)\,\in\, N_x(x,y)} |\nabla_x C_1(i,j) - \nabla_x C_2(i + d,j)| + \\ & \sum_{(i,j)\,\in\, N_y(x,y)} \left|\nabla_y C_1(i,j) - \nabla_y C_2(i + d,j)\right|. \end{aligned} \tag{7}$$

where N is a 5×5 window, $N_x(x, y)$is surrounding window without the right most columns and $N_y(x, y)$ is surrounding window without the lowest row. ∇_x is forward gradient to the right and ∇_y is forward gradient to the bottom. Color images are taken into account by summing up the dissimilarity measure for all channels. The left to right cross checking is used for occlusion handling and to find reliable correspondences. Winner takes all optimization is used to choose the disparity with lowest matching cost. The resulting dissimilarity measure is given by

$$C(x, y, d) = SAD\,(x, y, d) \; + \; Gradient(x, y, d). \tag{8}$$

2.4 Disparity Plane Fitting

After estimating the disparity map, instead of assigning disparity value to each pixel, the disparity plane is estimated by using disparities of each segment. The estimated disparity plane is assigned to each segment. A straight forward way to find the disparity plane parameters are by solving a linear square system. The commonly used least square methods are linear or median methods. We are used median solution to obtain the disparity plane.

3 Experimental Results

In order to evaluate the performance of the proposed stereo correspondence algorithm, we have carried out experiments using test images of stereo face database [22] and Middlebury stereo database [23]. The proposed algorithm is evaluated using a root mean square error and computation cost. The root mean square error and computation cost of the proposed method is compared with standard intensity based stereo matching methods.

3.1 Experiments on Stereo Face Images

To evaluate the performance of the proposed stereo correspondence algorithm, we have carried out experiments on test images of stereo face database [22]. It consists of stereo pairs of 70 subjects (35 males, 35 females) of size 676 × 516, recorded from six different viewpoints. The next five viewpoints range from a frontal to a profile view with respect to the viewing direction of the first camera, in equal steps of 1/8 radians. The intensities on the facial part of the image vary considerably over the different viewpoints.

The stereo images are first rectified using uncalibrated rectification to suppress the vertical displacement. Figure 1 shows the rectified stereo face images. The biorthogonal discrete multiwavelet transform is applied to rectified stereo images in order to decorrelate them into their subbands. The four quadrants contain approximation subband, horizontal detail subband, diagonal detail subband and vertical detail subband respectively. The two scale decomposition applied on left face image of Fig. 1 using discrete wavelet transform is shown in Fig. 2. Each subband size of one scale discrete wavelet transform is 169 × 129. The size of each subband of two scales is one by four of input image.

The approximation subband of two scale decomposition is used for color segmentation. The approximation subband contains low frequency components, and it won't have noise components. The result obtained after applying mean shift color segmentation method to the approximation subband is shown in Fig. 3a. The result of our method on face images is compared with the result of well known correlation based sum of absolute difference (SAD) algorithm. The disparity map resulted from our proposed method is shown in Fig. 3b and the Fig. 3c shows the disparity map produced by correlation based SAD algorithm. The results show that disparity map obtained by our algorithm is smoother than the result of correlation based SAD algorithm.

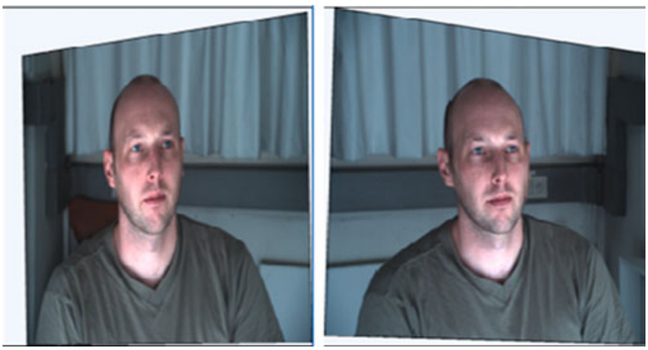


Fig. 1 Rectified *left* and *right* stereo images

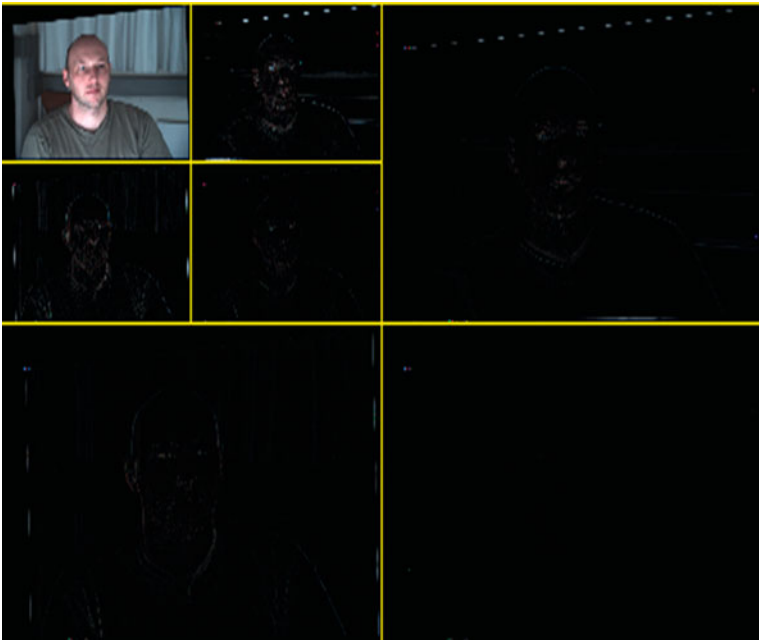


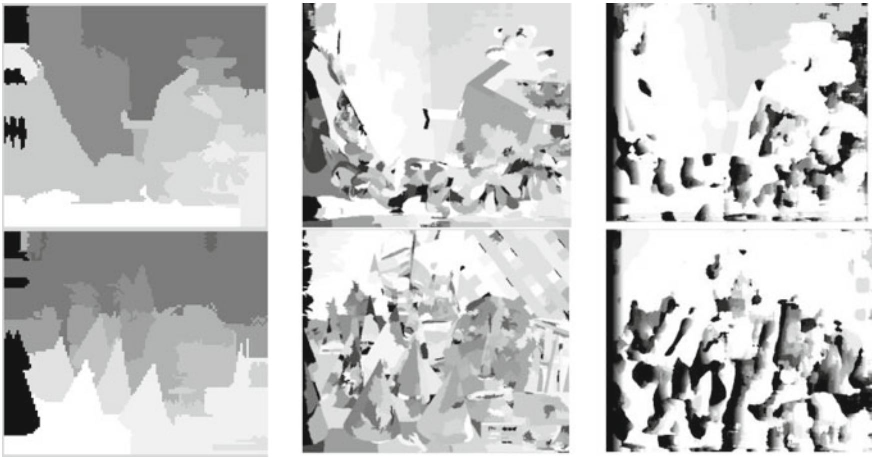
Fig. 2 Two scale Wavelet Decomposition of rectified left face image shown in Fig. 1

Fig. 3 **a** Homogeneous region obtained by color segmentation, **b** Disparity map of proposed method and **c** correlation based Sum of absolute difference

3.2 Experiments on Stereo Images of Middlebury Database

To validate the result of our algorithm with ground truth, experiments are performed on some of the stereo test images from the Middlebury stereo database [23]. We have used stereo images viz. Teddy, Cones and Aloe (first column of Fig. 6) of Middlebury database. Figure 4 shows the result of color segmentation applied on these images using a mean shift segmentation algorithm. The result of our method is compared with intensity based self adaptive dissimilarity and correlation based SAD algorithm. Figure 5 the shows estimated disparity map of these

Fig. 4 The results of mean shift segmentation



Proposed Method ISA Method CSAD Method

Fig. 5 Comparisons of estimated disparity map with intensity based self adaptive dissimilarity and correlation based SAD method

algorithms on stereo test images of Middlebury stereo database. The smooth disparity with less mismatch error is estimated by our algorithm when compared to other two methods in less computation time. To evaluate the proposed method, bad pixel map is estimated between the result of proposed method and corresponding image ground truth. The fourth column of Fig. 6 shows the bad pixel map. The bad pixel map result shows that mapping of bad pixel is very less.

Root mean square (RMS) method is a quantitative way to estimate the quality of the computed correspondences. The RMS error is calculated between the computed depth map $d_C(x, y)$ and the ground truth map $d_T(x, y)$, that is,

$$R = \left(\frac{1}{N} \sum_{(x,y)} |d_C(x, y) - d_T(x, y)|^2 \right)^{\frac{1}{2}}, \tag{9}$$

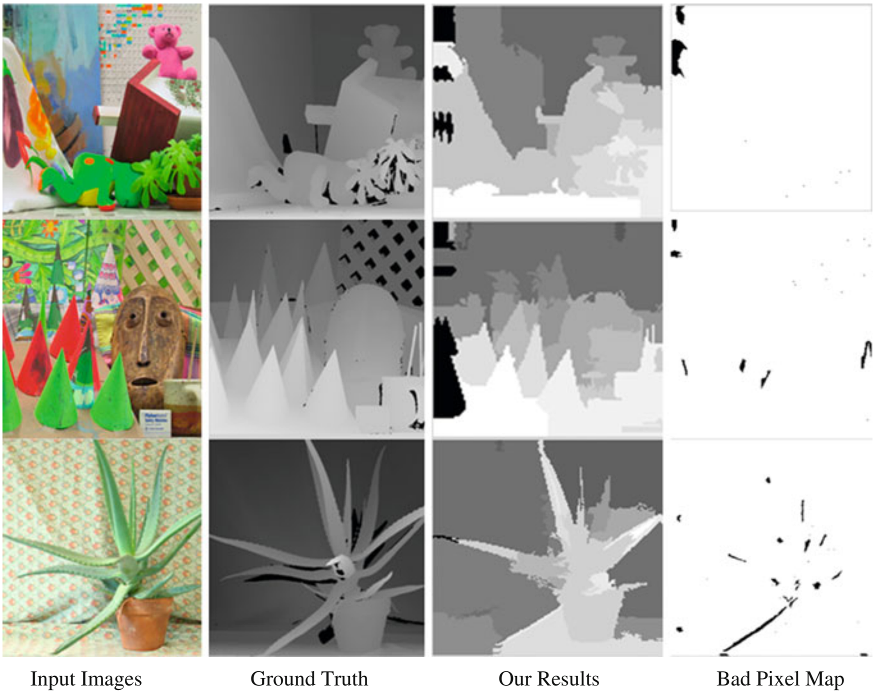


Fig. 6 Results on Middleburry datasets. From *top* to *bottom*: Teddy, Cones, aloe. From *left* to *right* reference images, ground truth disparities, the results of the proposed algorithm and the error images where the *black* region represents the erroneous pixels

where N is the total number of pixels. Table 1 shows the root mean square error estimated for results obtained by proposed method and correlation based SAD algorithm. The root mean square error of correlation based SAD algorithm is more than root mean square error of proposed method.

Table 2 compares computational time of proposed method, intensity based self-adaptability method (ISA) and correlation based SAD (CSAD) method for Face, Teddy, Cones and Aloe dataset using a PC with processor Intel core i3 with speed 2.27 GHz and 4 GB RAM. The proposed algorithm is implemented using MATLAB R2010a. The computational costs of proposed stereo correspondence method are reduced by 65–71 % in comparison with intensity based self adaptability method. Our proposed method generates a smooth disparity map in less computation time.

Table1 Root mean square errors

	Teddy	Cones	Aloe
Proposed method	0.0337	0.0628	0.0349
CSAD	0.0834	0.0922	0.0365

Table 2 Computational cost comparison of the proposed method and self adaptive dissimilarity method

Image pair	Image size	Execution time (m s)		
		CSAD	ISA	Proposed method
Face	676 × 516	32.69	24.87	4.14
Teddy	375 × 450	20.36	9.64	2.95
Cones	375 × 450	25.67	11.21	3.17
Aloe	555 × 641	43.58	29.94	10.34

4 Conclusions

In this paper, we have presented a new segment-based stereo matching technique using wavelets. In our algorithm, wavelet coefficients obtained by two scale DWT and relative changes in the color are used to estimate the disparity map. The reliable correspondences are obtained by using color segmentation and disparity plane fitting technique. The results obtained by our proposed method is compared with intensity based self-adaptability method, and correlation based SAD method. The root mean square error and computation time of proposed method is lesser than other two methods. The results show that the proposed technique produces a smooth disparity map with significantly fewer mismatch errors in less computation time.

References

1. Muhlmann K, Maier D, Hesser R, Manner R (2001) Calculating dense disparity maps from color stereo images, an efficient implementation. In: Proceedings of the IEEE workshop on stereo and multi-baseline vision (SMBV 2001), pp 30–36
2. Di Stefano L, Marchionni M, Mattoccia S, Neri G (2004) A fast area-based stereo matching algorithm. Image Vis Comput 22:983–1005
3. Yoon KJ, Kweon IS (2006) Adaptive support-weight approach for correspondence search. IEEE Trans Pattern Anal Mach Intell 28:650–656
4. Seok Y, Lee S (2011) Robust stereo matching using adaptive normalized cross correlation. IEEE Trans PAMI 33(4):807–822
5. Hamzah R, Hamid A, Md. Salim S (2010) The solution of stereo correspondence problem using block matching algorithm in stereo vision mobile robot. IICRD, pp 733–737
6. Bobick AF, Intille SS (1999) Large occlusion stereo. Int J Comput Vis 33(3):181–200
7. Kang SB, Szeliski R, Jinxjang C (2001) Handling occlusions in dense multi-view stereo. Proceedings of the IEEE conference computer vision and pattern recognition, vol 1, pp 103–110
8. Kim H, Yang S, Sohn K (2003) 3D reconstruction of stereo images for interaction between real and virtual worlds. In: Proceedings of the IEEE international conference on mixed and augmented reality
9. Ogale AS, Aloimonos Y (2008) Robust contrast invariant stereo correspondence. Proceedings of the IEEE international conference on robotics and automation, ICRA 2005, pp 819–824
10. Bleyer M, Gelautz M (2005) A layered stereo matching algorithm using image segmentation and global visibility constraints. ISPRS J Photogramm Remote Sens 59(3):128–150

11. Bleyer M, Gelautz M (2005) Graph-based surface reconstruction from stereo pairs using image segmentation. In: SPIE, vol 5665, pp 288–299
12. Deng Y, Yang Q, Lin X, Tang X (2005) A symmetric patch-based correspondence model for occlusion handling. In: ICCV, pp II:1316–1322
13. Hong L, Chen G (2004) Segment-based stereo matching using graph cuts. In: CVPR, vol I, pp 74–81
14. Xiao J, Xia L, Lin L (2010) A segment based stereo matching method with ground control points. In: IEEE transaction on ESIAT
15. Mallat S (1999) A wavelet tour of signal processing. Academic Press, New York
16. Sarkar I, Bansal M (2007) A wavelet-based multiresolution approach to solve the stereo correspondence problem using mutual information. IEEE Trans Syst Man Cybern 37:1009–1014
17. Begheri P, Sedan CV (2010) Stereo correspondence matching using multiwavelets. In: Fifth international conference on digital telecommunication
18. Bhatti A, nahavandi S, Hossny M (2010) Wavelets/Multiwavelets bases and correspondence estimation problem; an analytic study. In: 11th international conference on control, automation, robotics and vision
19. Klaus A, Sormann M, Karner K (2006) Segment-based stereo matching using belief propagation and a self-adapting dissimilarity measure. In: Proceeding of the ICPR
20. Comaniciu D, Meer P (2002) Mean shift a robust approach toward feature space analysis. IEEE PAMI 24:603–619
21. Fusiello A, Irsara L (2008) Quasi-euclidean uncalibrated epipolar rectification. In: ICPR, pp 1–4
22. http://cvlab.epfl.ch/data/strechamVS/ (2010)
23. Middlebury database (2010) http://vision.middlebury.edu/stereo/data/

Automated Calibration of Microscope Based on Image Processing Methods

N. B. Byju and R. Rajesh Kumar

Abstract Microscopes must be calibrated so that accurate measurements can be made. Pixel size calibration is necessary for medical image processing applications. To calibrate a digital microscope, the normal procedure is to manually measure the distance of the divisions in the image of stage micrometer and find out the pixel size calibration factor. Here we propose an automated methodology for finding pixel size calibration factor of the digital microscope by analyzing the microscopic image of stage micrometer. The proposed methodology can reduce the human observational errors in manual method of calibration. Our approach is scalable and well suited for automated image acquisition applications.

Keywords Calibration · Digital microscopes · Image processing · Stage micrometer · Segmentation

1 Introduction

Since its invention, the microscope has been a valuable tool in the development of scientific theory. The function of any microscope is to enhance resolution. The microscope is used to create an enlarged view of an object at different magnification levels such that we can observe details not otherwise possible with the human eye. The microscope consists of different objectives, 10×, 20×, 40× etc. to

N. B. Byju (✉) · R. Rajesh Kumar
Centre for Development of Advanced Computing, Trivandrum, India
e-mail: byju.nb@gmail.com

R. Rajesh Kumar
e-mail: rajeshkumar@cdac.in

Mohan S. and S. S. Kumar (eds.), *Proceedings of the Fourth International Conference on Signal and Image Processing 2012 (ICSIP 2012)*, Lecture Notes in Electrical Engineering 222,
DOI: 10.1007/978-81-322-1000-9_9,

obtain different magnification levels A microscope can be used not only to see very small objects but also to accurately measure them [1]. A convenient unit for microscopic specimen is micrometers (μm). For conventional microscopic usage, i.e. when viewing the specimen through the eye-piece, the microscopist can make measurements through a calibrated eye-piece reticle [2]. The reticle is calibrated using a stage micrometer [3]. Presented in the Fig. 1 is a stage micrometer, which contains a small metallized millimeter ruler that is subdivided into increments of 10 and 100 μm. Juxtaposing the graduations on the eyepiece reticle with those on the stage micrometer enables the microscopist to calibrate the reticle gauge and perform linear measurements on specimens. When calibrating, we need to line up the stage micrometer with the ocular scale and count the number of divisions on the ocular scale per millimeter or micrometer on the staged micrometer. This will give the distance in micrometers per ocular unit. The number of divisions will change as the magnification changes. The conversion to other magnifications is accomplished by factoring in the difference in magnification.

In a digital microscope, a digital camera (CCD or CMOS) replaces the eyepiece to output a digital image to a monitor. The camera used for image capturing also supports different resolutions, typically in the range 1.4–12 MP. With a typical 8 MP camera, a 3264 × 2448 pixels image is generated. The resolution of the image depends on the field of view of the lens used with the camera. For the image processing applications, it is important to know the conversion formula from pixels to physical units (i.e. how many pixels make 1 μm) [4]. This is can be achieved by calibrating the microscope for each different objective setting [1]. Even though the absolute measurement of calibration is possible from the specifications of the microscope and camera, the tolerances in the sensor and the imperfections in the stage micrometer may change the absolute value. Hence digital image of stage micrometer can be used for calibrating digital microscope. The digital images of stage micrometer at 10 × and 40 × magnification are given in Fig. 2.

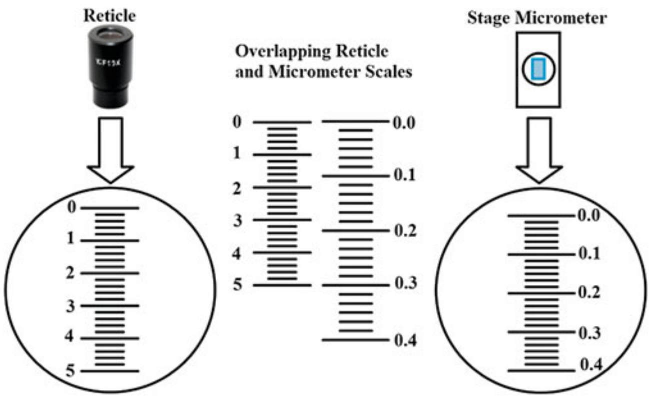


Fig. 1 Stage micrometer and eyepiece reticle

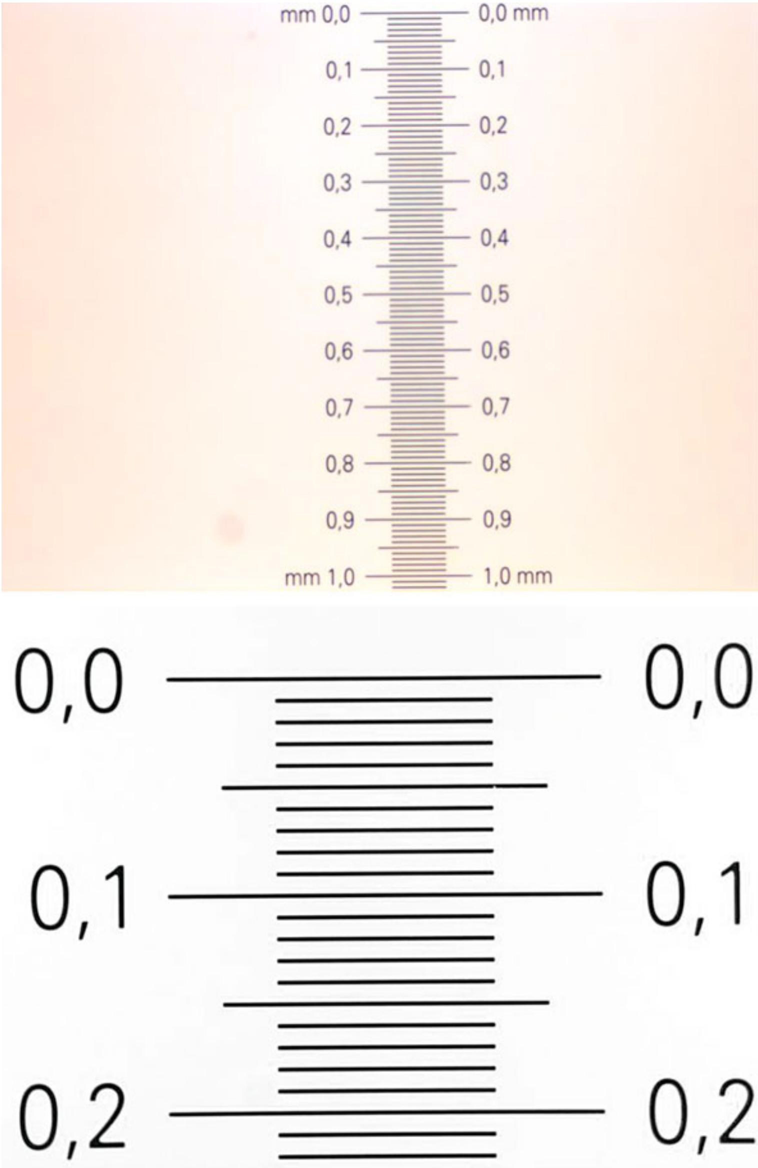
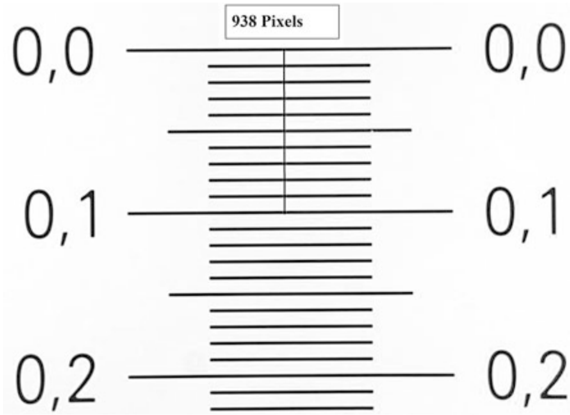


Fig. 2 Digital images of stage micrometer at 10× (*above*) and at 40× (*below*) magnification

In order to calibrate digital microscope, image processing methodology [5] is proposed in this work. In the rest of the paper we first define the problem. We then describe the existing manual method of calibration. Next describes the proposed image processing methodology. The experiments and results describe the

Fig. 3 Manual selection of pixels for calibration factor



calibration details at various magnifications and compares manual with automatic calibration [6]. Next section concludes the article.

2 Manual Method of Calibration

For a digital microscope, where the camera output a digital image to a monitor, may not have the provision to observe the sample directly through an eyepiece. In this scenario, calibrating the microscope is required through the analysis of the image of stage micrometer. The manual calibration method of digital microscope is to manually measure the number of pixels between certain divisions in stage micrometer using specific software tools to calculate the calibration factor. The manual selection is given in Fig. 3. Here number of pixels are measured from the graduation 0.0–0.1 mm and is found to be 938 pixels for the given specimen.

The calibration factor is defined as,

$$\text{Calibration Factor } = \frac{\mathrm{T_d} * 1}{\mathrm{P}}$$

where, T_d is total number of divisions in selected in stage micrometer image, l is length of one division in micrometer and P is the total number of pixels constituting the selected divisions. For the given example,

Calibration Factor = 10 * 0.01/938 = 0.106 μm.

3 Problem Definition

An important consideration when using a stage micrometer image to calibrate a digital microscope is to include as many of the stage micrometer graduations as possible in the calculation [7, 8]. This will minimize errors due to variations in the individual graduation intervals, in addition to the potential error in precisely identifying the edges of individual lines. Another important factor that should be considered as a potential source of measurement error is the subjectivity involved in setting a reference line at the edge of a division line [9]. In manual calibration, measurement error may be high due to the subjectivity in determining the edge of graduation intervals. The proposed approach aims at eliminating the errors of manual calibration by automating the method of calibration through image processing methodology. The basic assumption is that all pixels are squares and there is no image distortion in any direction.

4 Methodology

Figure 4 gives the block diagram of the proposed automatic calibration of microscopic image. In the initial stage the calibration slide is digitized at various magnifications. Next the image is preprocessed using an edge preserving median filter to remove the noise without affecting edges [10]. The segmentation module uses gradient detection of the image using Laplacian of Gaussian filter with a threshold of 0.11 to detect the binary image objects. Small objects are removed from the segmented objects to retain only significant objects, in this case, lines and letters.

Feature extraction module is used to analyze the features of the segmented objects in order to retain only lines. One of the features used is the width of the objects, since line segments will have more width compared to the other objects present in the image. The other feature used is the elongation of the objects which

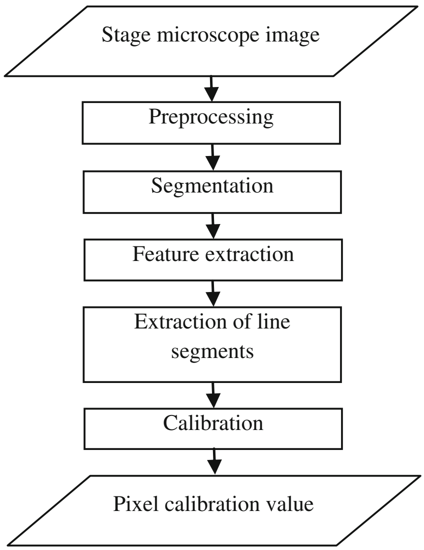


Fig. 4 Block level description of the proposed methodology

are in the range 0–1 [10]. Moments of the objects are used for calculating the elongation.

Central moments of the objects can be defined as,

$$\mu_{pq} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x-\bar{x})^p (y-\bar{y})^q f(x, y)\, dx\, dy$$

where $\bar{x}$ and $\bar{y}$ are the components of the centroid. If *f*(x, y) is a digital image, then the previous equation becomes,

$$\mu_{pq} = \sum_x \sum_y (x-\bar{x})^p (y-\bar{y})^q f(x, y)$$

The covariance matrix of the image *f*(x, y) is now

$$\mathrm{COV}[f(x,y)] = \begin{bmatrix} \mu'_{20} & \mu'_{11} \\ \mu'_{11} & \mu'_{20} \end{bmatrix}$$

where,

$$\mu'_{20} = \mu_{20}/\mu_{00}$$
$$\mu'_{02} = \mu_{02}/\mu_{00}$$
$$\mu'_{11} = \mu_{11}/\mu_{00}$$

The Eigen values of the covariance matrix can be calculated as,

$$\lambda_i = \frac{\mu'_{20} + \mu'_{02}}{2} \pm \frac{\sqrt{4\mu'^2_{11} + (\mu'_{20} - \mu'_{02})^2}}{2}$$

The relative difference in magnitude of the Eigen values are thus an indication of the eccentricity of the object, or how elongated it is. Then the eccentricity of the object can be calculated as

Table 1 Elongation (Eccentricity) of letters and lines

Objects	Eccentricity
Lines	0.9998
m	0.7560
,	0.9791
0	0.7634
1	0.9815
2	0.9159
3	0.8757
4	0.7965
5	0.8644
6	0.8063
7	0.9144
8	0.8106
9	0.8062

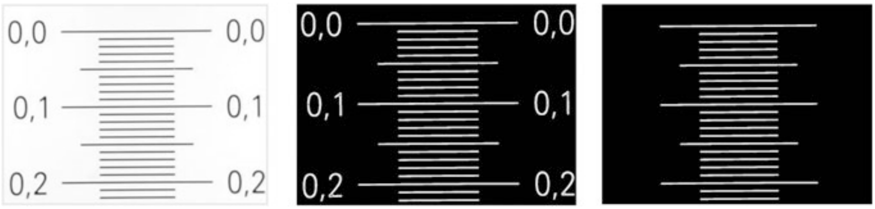


Fig. 5 Extraction of *lines*; *left*: input image, middle: segmented image, *right*: *lines* extracted image

Fig. 6 Portion of the line segmented stage micrometer binary image after thinning and horizontal stretching

$$1 - \frac{\lambda_2}{\lambda_1}$$

The value is between 0 and 1. An eccentricity of 0 means the object is a circle and a long, thin object (line segment) has an eccentricity that approaches 1. The Table 1 shows the elongation values of letters and lines.

Stages of extraction of the divisions in stage micrometer image are shown in Fig. 5. It shows the input image, segmentation of the input image and the extraction of lines after removal of letters and other objects.

The calibration module applies thinning and horizontal stretching of lines extracted to obtain central lines of line segments for measuring calibration factor. The lines extracted and super imposing of the central lines obtained on the input image are shown in Fig. 6. The first column vector of the binary image is then used for calculating calibration factor. In order to minimize calibration error, the method uses all the graduation lines present in the image for applying the calibration formula.

It is observed from the above Fig. 7 that, the divisions are not exactly horizontal, rather having an angle $\theta°$ difference with horizontal axes. It may be because of the misalignment of camera sensor or the stage micrometer is not exactly placed

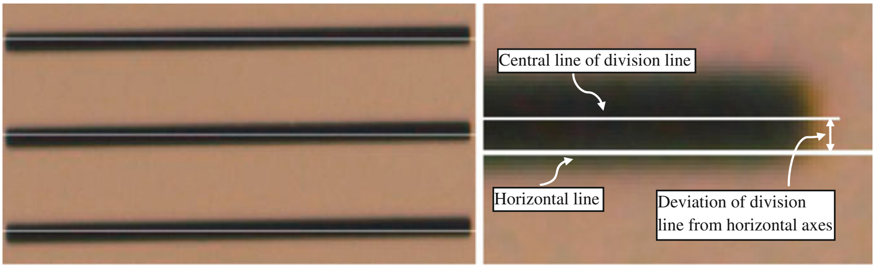


Fig. 7 *Left*: Thin lines extracted super imposed on the original image, *right*: Tilt of stage micrometer from horizontal

Fig. 8 Angle shift of stage micrometer divisions with horizontal axes

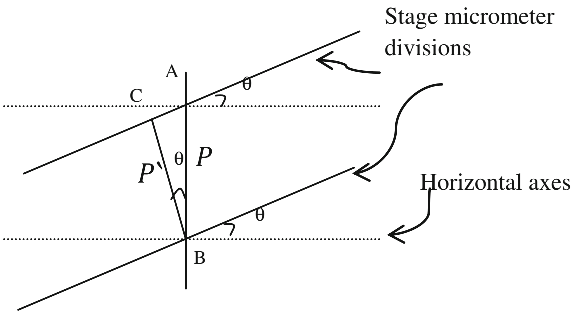
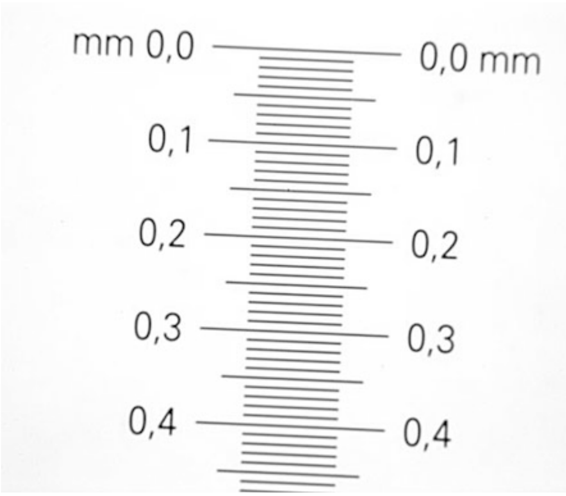


Fig. 9 Stage micrometer image having an angle shift of 2.5°



horizontally. Hence the value of P which is the total number of pixels from first division to last division may affected by an angle θ. This can be corrected by measuring the new value of P as given in Fig. 8.

Table 2 Measurements taken using the proposed method

Magnification Levels	T_d	P' (pixels)	l(μm)	Calibration Factor (μm)
10×	102	2387	0.01	0.4273
20×	51	2395	0.01	0.2129
40×	22	2065	0.01	0.1065

Hence the new value of P

$$\mathrm{P}' = \mathrm{P}\cos\theta,$$

Here in our experiments, the divisions in the stage micrometer are having an angle shift of 0.4° with the horizontal axes. Another microscopic digital image of stage micrometer having an angle shift of 2.5°, which causes 0.1 % error in calibration, is shown in Fig. 9.

5 Experiments and Results

The dataset consists of digitized images of stage micrometer at 10×, 20×and 40 magnification. The images are taken at 8 MP resolution using LEICA DFC495 camera attached to LEICA DM2500 microscope. Two microscopes are used to generate the data set. CCD is selected over CMOS due to higher quality and lower noise [11]. Sensor size is 8.8 × 6.61 mm (Type 2/3); video adaptor 0.63 × de-magnification and pixel size is 2.7 × 2.7 μm. Each image is auto calibrated for pixels using the proposed method and the results are obtained for calibration. The improvement in error reduction compared to manual calibration is also analyzed from the results obtained. The proposed methodology identifies the central line of the line segment to measure distance in terms of pixels from first line to the last line. The measurements taken for the proposed method over various magnification levels are given in Table 2.

The calibration values from the data sheet specification, manual calibration method and the proposed automatic calibration method are given in Table 3. The absolute calibration value from data sheet specifications can be calculated as follows,

Absolute calibration Factor = pixel size/demagnification/magnification

i.e., at 40×,

Table 3 Calibration values using different methods

Calibration value	10×	20×	40×
Absolute calibration	0.4285	0.2142	0.1071
Manual method	0.4255	0.2119	0.1055
Automated method	0.4273	0.2129	0.1065

Table 4 Calibration error

Magnification	Manual calibration using microscope1 (tilt = 0.4°) (%)	Manual calibration using microscope2 (tilt = 2.5°) (%)
10×	0.42	0.53
20×	0.46	0.61
40×	0.96	1.03

Absolute calibration Factor = 2.7/0.63/40 = 0.1071 μm.

In the manual method of calibration, the subjectivity of determining the central line of lines of division will lead to the measurement errors. The calibration error may severely affect medical image processing applications which are very sensitive to size and shape based features. In the proposed method, since it identifies the central edges of division lines, there will not be any measurement errors. The error which can occur is the quantization error, which will occur only if the division lines have an even number of pixels constituting its height. The maximum quantization error will be 1 pixel irrespective of the magnification level. The measurement error of manual calibration against proposed automated method using the two different microscopes is given in Table 4. It is observed from the calculations, that at higher magnifications the pixel calibration error will be more for manual calibration. The manual calibration method is also affected by the horizontal orientation of stage micrometer, which is taken care in the proposed method.

6 Conclusion

The importance of image based automatic pixel calibration of Microscopic images is discussed in this paper. From the results it is evident that this methodology is well suited for automation of image acquisition applications where the measurements of the objects are considered in the image analyzing application. It can eliminate the measurement error induced by human intercepts in calibrating digital microscope. The proposed method is also eliminates the error introduced by the orientation of stage micrometer where the accuracy of manual method is affected by it. This approach is scalable to use any stage micrometer with minor modifications to suit different type of gradations and in stage micrometers.

References

1. Altinay D, Bradley AP, Mehnert A (2010) On the estimation of extrinsic and intrinsic parameters of optical microscope calibration. IEEE, digital image computing: techniques and applications, 2010, pp 190–195

2. Atthi N, Meerot S, Keawpoonsuk A, Jeamsaksiri W, Hruanun C, Poya A (2008) Fabrication of eyepieces lenses scale for optical microscope. IEEE proceedings of ECTI-CON 2008, pp 525–528
3. Penzes WB, Allen RA, Cresswell MW, Linholm LW, Teague EC (2000) A new method to measure the distance between graduation lines on graduated scales. IEEE Trans Instrum Meas 49(6):1285–1288
4. Figl M, Birkfellner W, Hummel J, Ede C, Hanel R, Bergmann H (2003) Calibration of an optical see through head mounted display with variable zoom and focus for applications in computer assisted interventions. In: Proceedings of SPIE: medical imaging 2003: visualization, image-guided procedures, and display, vol 5029, pp 618–623
5. Gonzalez R, Woods R (1992) Digital image processing. Addison-Wesley, Reading
6. Ammi M, Fremont V, Ferreira A (2009) Automatic camera based microscope calibration for a telemicro manipulation system using a virtual pattern. IEEE Trans Robot 25(1):184–191
7. Ralis SJ, Vikramaditya B, Nelson BJ (2000) Micropositioning of a weakly calibrated microassembly system using coarse-to-fine visual serving strategies. IEEE Trans Electron Packag Manuf 23(2):123–131
8. Boddeke FR, Van Vliet LJ, Young IT (1997) Calibartion of the automated z-axis of a microscope using focus functions. J Microsc 186(Pt 3):270–274
9. Chavez G, Valera B, Padilla S (2004) Automated calibration of graduated scales. 1st international conference on electrical and electronics engineering, pp 385–390
10. Rajesh Kumar R, Ajith Kumar V, Sharath Kumar PN, Sudhamony S, Ravindrakumar R (2011) Detection and removal of artifacts in cervical cytology images using support vector machine. In: Proceedings of IEEE symposium on IT in medicine and education (ITME), Dec 2011, pp 717–721, doi: 10.1109/ITiME.2011.6130760
11. Theuwissen AJP (2001) CCD or CMOS image sensors for consumer digital still photography? IEEE, VLSI technology, systems, and applications, 2001, pp 168–171

Classification of Handwritten Document Image into Text and Non-Text Regions

V. Vidya, T. R. Indhu and V. K. Bhadran

Abstract Segmentation of document image into text and non-text regions is an essential process in document layout analysis which is one of the preprocessing steps in optical character recognition. Usually handwritten documents has no specific layout. It may contain non text regions such as diagrams, graphics, tables etc. In this work we propose a novel approach to segment text and non text components in Malayalam handwritten document image using Simplified Fuzzy ARTMAP (SFAM) classifier. Binarized document image is dilated horizontally and vertically and merged together. Perform connected component labelling on the smeared image. A set of geometrical and statistical features are extracted from each component and given to SFAM for classifying it into text and non text components. Experimental results are promising and it can be extended to other scripts also.

Keywords Text and not text regions segmentation · Simplified fuzzy ARTMAP

V. Vidya (✉) · T. R. Indhu · V. K. Bhadran
Centre for Development of Advanced Computing, Thriuvananthapuram, Kerala, India
e-mail: vidyav@cdac.in

T. R. Indhu
e-mail: indhu@cdac.in

V. K. Bhadran
e-mail: bhadran@cdac.in

Mohan S. and S. S. Kumar (eds.), *Proceedings of the Fourth International Conference on Signal and Image Processing 2012 (ICSIP 2012)*, Lecture Notes in Electrical Engineering 222,
DOI: 10.1007/978-81-322-1000-9_10,

1 Introduction

Recently document image analysis has become very prominent in the field of computer applications. Wide variety of information, which has been conventionally stored on paper, is now being converted into electronic form for better storage and intelligent processing. The purpose of document image analysis is to identify the text and non text components in image of document and to extract intended information from them.

Exclusion of non text components from printed/handwritten document image is an important pre-processing step in optical character recognition system. Automatic exam paper evaluation, digitization of notebooks, form processing etc. are major application areas of offline handwritten recognition. These kinds of documents have large number of non-text components such as diagrams, tables, graphics etc. Many methods have been proposed to deal with a wide variety of machine-printed documents [1–5] whereas only a few methods were dedicated to handwritten documents.

In general, the problem of classifying text and non-text components in document image is grouped into three categories (i) block or zone based classification (ii) pixel based classification (iii) component based classification. In zone based approach entire document is segmented into zones and then classify the obtained blocks into a set of determined classes. In [1], area Voronoi diagram is used to segment the document into number of zones and extract structural and textural image zone features. Partial Least Squares method is applied on it to compute pair wise discriminating features and given to a hybrid classifier. Keysers et al. [2] proposed a run length histogram based features extraction and nearest neighbor classifier to segment text and non text component.

Pixel based approach classifies individual pixels instead of zones according to predefined classes. Moll et al. [3] classify individual pixels instead of regions, to avoid the constriction of the limited classes of region shapes. The approach is applied on handwritten, machine printed and photographed document images. A color-based representation of document images using pixel-correspondence graph is described by Shafait et al. [4]. Pixel based classification approaches are slow with respect to execution time.

Instead of pixel or zone, component based approach, classifies each connected component into text and non-text component. Bukhari et al. [5] proposed a self-tunable multilayer perceptron (MLP) classifier for distinguishing between text and non-text connected components using shape and context information as a feature vector.

In addition to these three methods, multi resolution morphological based method is proposed by Bloomberg et al. [6] for text and half tone components. This approach comprises three steps: (1) in first step, seed image is generated by sub sampling input image such that the resulting seed image mainly contains halftone pixels. (2) Then mask image is produced by using morphological operations. (3) in last step binary filling operation is used to transform seed image with the help of mask image into final halftone mask image. Bukhari [7] introduced improvements in Bloomberg's text/image segmentation algorithm to generalize it for separating text and non-text components including halftones, drawings, graphs etc.

Handwritten documents usually have no specific layout due to its unstructured data. Ram Sarkar et al. [8] proposed a modified Run Length Smearing Algorithm (RLSA), called Spiral RLSA (SRLSA), to suppress the non-text components from text ones in handwritten document images. In this method, document pages are classified into text/non-text groups using a Support Vector Machine (SVM) classifier.

Here we present a component based segmentation approach to separate text and non text regions in a handwritten document image using simplified Fuzzy ARTMAP (SFAM) Classifier. ARTMAP is a competitive neural network which is capable of learning many new things without forgetting things learned in past. After classification non text regions are removed from the document and the non text free-document image is fed to offline handwritten recognition (OHR) system. We have experimented with Malayalam script document images.

2 Handwritten Document Image Segmentation

In our work, document image segmentation process is divided into three stages such as preprocessing, feature extraction and classification. Preprocessing refines the input image in such a way that makes it suitable for extracting features. Steps involved in preprocessing are binarization, noise reduction, vertical and horizontal dilation and connected component labelling of the joined image. Instead of taking components directly from binarized image as in [5, 8], we used dilated image for extracting components which makes it easier to identify non text components. In feature extraction stage, a set of geometrical and statistical features are extracted from each connected component. This feature vector is fed to SFAM for classification of components into text and non-text.

2.1 Preprocessing

Let $I(x, y)$ be H $\times$ W sized gray scale image. Using Otsu' binarization Technique [9], threshold T is computed and gray scale image is converted to binary image $B(x, y)$ using Eq. 1.

$$B(x,y) = \begin{cases} 0 & I(x,y) <= T \\ 1 & I(x,y) > T \end{cases} \tag{1}$$

Modified Directional Morphological Filter (MDMF) algorithm [10] removes the noise present in the image. MDMF consists of four erosion sub-elements R_1,

R_2, R_3, R_4 to erode irregular noises in horizontal, vertical and two diagonal directions followed by one dilation operation R5 as per (2), for a restorative effect of linking broken strokes and remedying eroded pixels in the character strokes.

$$B' = ((B\Theta R_1) \cup (B\Theta R_2) \cup (B\Theta R_3) \cup (B\Theta R_4)) \oplus R_5 \quad (2)$$

Text line gap estimated from vertical projection profile of the image *B'(x, y)* is used as the height of the vertical structuring element, V and half of the average character width as the width of V. The image *B'(x, y)* is dilated vertically as given in (3) so that lines in the image join together.

$$BV = B' \oplus V \quad (3)$$

The image *B'(x, y)* is horizontally dilated by (4) so that words in a text line touch each other. From a small horizontal strip of image the horizontal projection profile is computed. Gap between the words obtained from the profile is assigned as the width of horizontal structuring element *H* and half of the average character height as the height of *H*.

$$BH = B' \oplus H \quad (4)$$

The superimposed image, *J(x, y)* is obtained by merging two images BV and BH as in

$$J(x, y) = BH \& BV \quad (5)$$

Holes in the image are filled by applying morphological closing operation. Original image and superimposed image are shown in Fig. 1. All the connected components in the superimposed image *J(x, y)* are identified by connected component labelling [11]. Each connected component is given to feature extraction module.

2.2 Feature Extraction

All components in the labelled image are not required for extracting features. Components with size lesser than a threshold are assigned as text since non-text components have a minimum size. Components with size greater than a threshold are used for extracting the features. Using bounding box information of connected components, *B'(x, y)* is cropped and features are extracted from the image thus obtained. Let, H_{Ci}, W_{Ci} be the height and width of the ith component C_i., A_{BCi}, A_{JCi} are the area of the component in binarized image, superimposed image respectively. The area is defined to be the number of foreground pixels belonging to the region. For each component Ci, calculate the features such as:

- Height ratio: H_{Ci}/H
- Width ratio: W_{Ci}/W

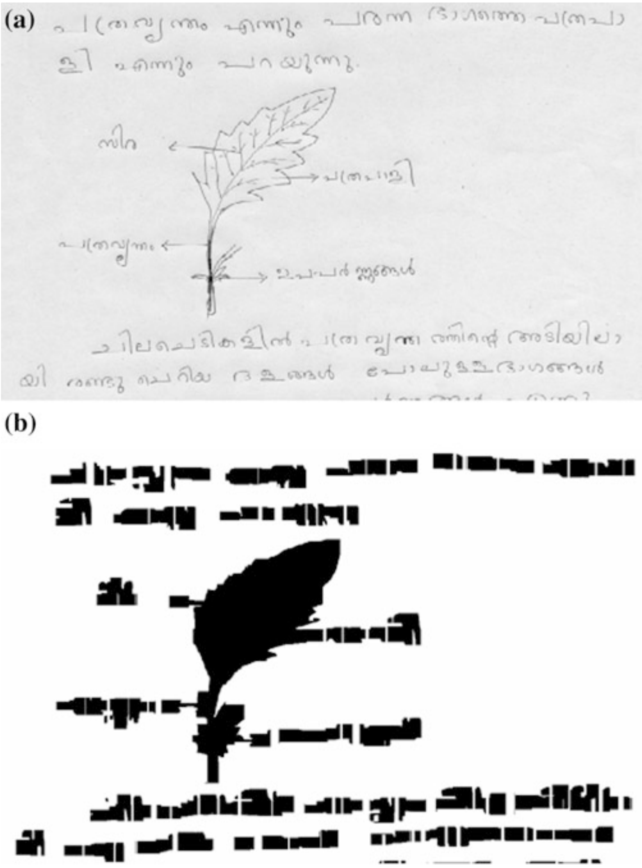


Fig. 1 **a** Original image **b** Superimposed image after fill holes

- Area ratio in combined image: A_{JCi}/total area of image J.
- Area ratio in binarized image: A_{BCi}/total area of image B'.
- Pixel density in combined image: $A_{JCi}/(H_{Ci} \text{ X } W_{Ci})$
- Pixel density in original image: $A_{BCi}/(H_{Ci} \text{ X } W_{Ci})$
- Mean and standard deviation of normalized vertical histogram
- Mean and standard deviation of normalized horizontal histogram
- Mean and standard deviation of the component image in original image
- Gap flag: If number of zeros in the vertical projection profile is greater than a threshold, set it to 1 otherwise 0.
- Large component flag: Analyze each element of component in the original image. If it contains the large sized element as almost the size of the component then flag is set to 1 otherwise 0.

The resultant feature vectors are fed to SFAM for training and classification of text and non-text components.

2.3 Simplified Fuzzy ARTMAP Classifier

Fuzzy ARTMAP refers to a neural network architecture based on Adaptive Resonance Theory (ART) that is capable of fast, stable, on-line, unsupervised or supervised, incremental learning, classification, and prediction in response to arbitrary sequences of analogue or binary input vectors [12]. Simplified Fuzzy ARTMAP (SFAM) is simpler and faster version of Fuzzy ARTMAP (FAM) by removing redundancies [13]. Simplified ARTMAP architecture shown in Fig. 2.

Fuzzy ART neural network consists of two fully connected layers of nodes: input layer F_1 with M nodes, and competitive layer, F_2 with N nodes. F_1-to-F_2 layer connections associated with a set of real-valued weights $W = \{w_{ij} \in [0, 1]$ where $i = 1,2,\ldots,M$; $j = 1,2,\ldots N\}$.Each F2 node j represents a recognition category that learns a prototype vector $w_j = (w_{ij}, w_{2j}, \ldots, w_{Mj})$. Through learned associative links, F_2 layer of fuzzy ART is connected to an L node map field F^{ab}, where L is the number of classes in the output space. Learning rate ($\beta \in [0, 1]$), choice ($\alpha > 0$), match tracking ($0 < \varepsilon < 1$) and baseline vigilance ($\rho \in [0, 1]$) are the parameters of SFAM. Steps for SFAM are:

1. **Initialization:** All the weight values w_{ij} associated with the F_1–to–F_2 layer and w_{ij}^{ab} associated with the F_2–to–F^{ab} connections are initialized to 1. Assign values to all the SFAM parameters.
2. **Input pattern coding:** When a training pair (a,t) is presented to the network, input pattern 'a' is complemented coded such as $A = (a, a^c)$, where $a_i^c = (1 - a_i)$ and $a_i \in [0,1]$. The vigilance parameter ρ is reset to its baseline value $\bar{\rho}$.
3. **Prototype selection:** Pattern A activates layer F_1 and is propagated through weighted connections W to layer F_2. Activation of each node j in the F_2 layer is determined by the choice function:

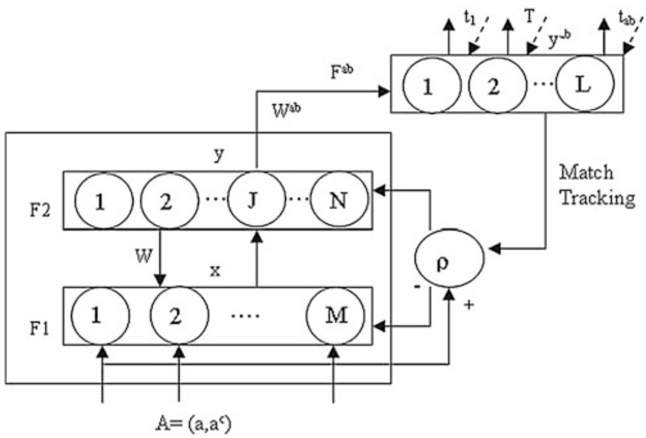


Fig..2 Simplified fuzzy ARTMAP architecture

$$T_j(A) = \frac{|A \wedge w_j|}{\alpha + |w_j|} \tag{6}$$

where | | is the norm operator, ^ is the fuzzy AND operator. The node which has the greatest activation value remains active. Choose arbitrary node if more than one T_j is maximal. The winner Node J propagates its top-down vector w_J back onto F_1 and vigilance test is performed. This test compares the degree of match between w_J and A against the dimensionless vigilance parameter ρ:

$$\frac{|A \wedge w_j|}{|A|} = \frac{|A \wedge w_j|}{M} \geq \rho \tag{7}$$

If the test is passed, then node J remains active and resonance is said to occur. Otherwise, the network inhibits the active F_2 node and searches for another node J that passes the vigilance test. If such a node does not exist, an uncommitted F_2 node becomes active and undergoes learning (Step 5).

4. **Class Prediction:** Pattern t is fed directly to the Map Field F^{ab}, while the $F2$ category y learns to activate the map field via associative weights W^{ab}. The F^{ab} layer produces a binary pattern of activity $(y^{ab} = y_1^{ab}, y_2^{ab}, \ldots \ldots y_L^{ab}) = t \wedge w_j^{ab}$ (in which the most active F^{ab} node K yields the class prediction *(K = k (J))*. If node K constitutes an incorrect class prediction, then a match tracking signal raises the vigilance parameter ρ just enough:

$$\rho = \frac{|A \wedge w_j|}{M} + \varepsilon \tag{8}$$

   where $\varepsilon = 0+$, to induce another search among F_2 nodes in Step 3. This search continues until either an uncommitted F node becomes active (or learning directly ensues in Step 5) or a node J that has previously learned the correct class prediction K becomes active.

5. **Learning:** Learning input involves updating vector w_j, and if J corresponds to a newly-committed node, creating an associative link to F^{ab}. The prototype vector of F_2 node J is updated according to:

$$W_J' = \beta(A \wedge w_J) + (1 - \beta)w_J \tag{9}$$

   where β is a fixed learning rate parameter and we set algorithm to fast learning with $\beta = 1$. A new association between F_2 node J and F^{ab} node K is learned by setting $W_{ij}^{ab} = 1$ for $K = k$, where K is the target class label for input pattern and 0 otherwise. The next training subset pair *(a, t)* is presented to the network in Step 2.

Table 1 Number of components used for training and testing

	No. of Training samples	No. of Testing samples
Text	116	152
Non text	35	43
Total	151	195
Accuracy (%)	100	94.87

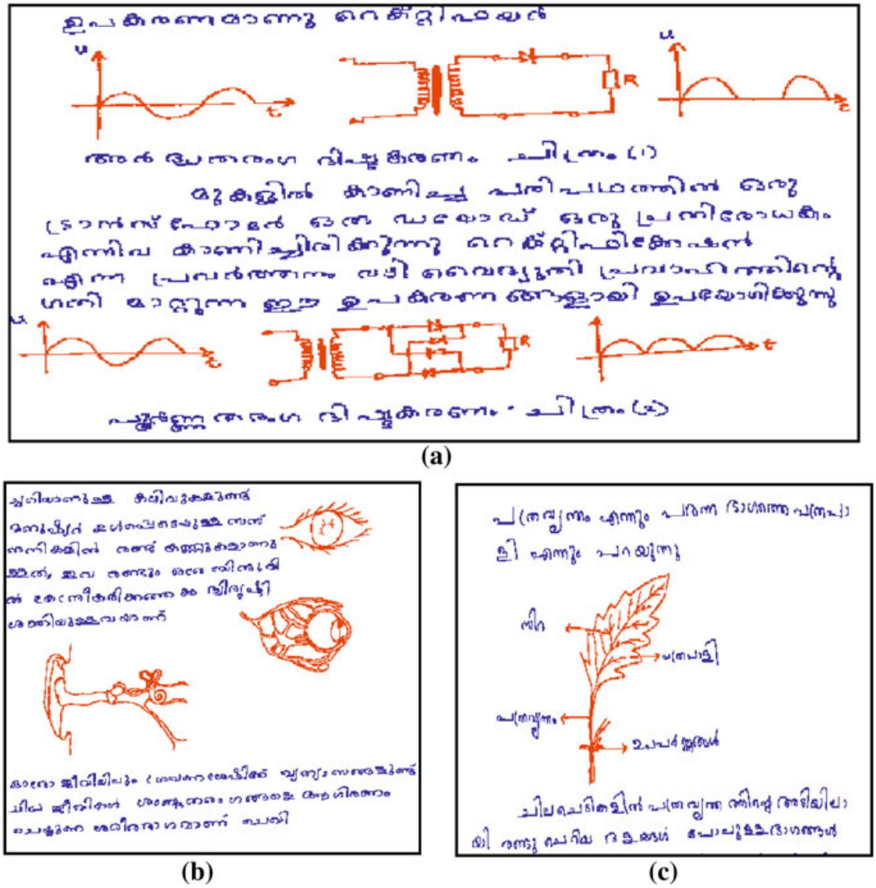


Fig. 3 Output images after classifying text and non text components. *Red* color indicates non text and *blue* color is text component

3 Experimental Results

We have experimented on a small set of handwritten document images. 151 components from 14 images were given to SFAM classifier for training. The parameters used are $\rho = 0.9$, $\alpha = 0.001$ and $\beta = 1$. 195 components from 16

Table 2 Classification rates

Component	Classification rate in %	
	Text	Non text
Text	97.38	3.85
Non text	2.62	96.15

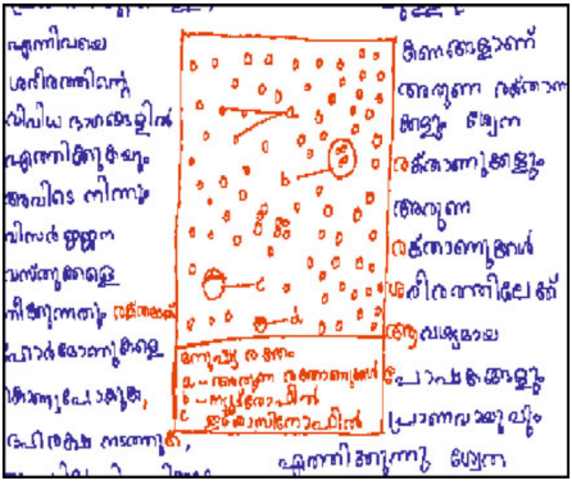


Fig. 4 Segmentation error

images were used for testing. Distribution of text and non text components used for training and testing and the accuracies achieved are given in Table 1.

Segmented images are shown in Fig. 3, in which text components are in blue color and non-text components in red color. In our experiments, text components coming within non text component are classified as non-text components. Text and non text classification and its misclassification rates are shown in Table 2.

The main issue we noticed in our work was text components adjoined to non text components gets misclassified as non-text components as shown in Fig. 4.

4 Conclusions

Classification of text and non text components in handwritten document image is presented in this paper. Here we have mainly focused on non text components such as drawings, diagrams, tables and graphics. Suitable features are taken from components of superimposed image and given to simplified Fuzzy ARTMAP classifier. An average segmentation accuracy of 97.11 % is obtained for Malayalam handwritten document images. Incremental property of this classifier can be used to train new type of non-text components such as symbols, mathematical notations, rulers etc. This method can be extended to other languages also. Our future work aims to further improve the current segmentation approach by separating adjoined text and non text components and also to test more number of document images.

References

1. Abd-Almageed W, Agrawal M, Seo W, David D (2008) Document-zone classification using partial least squares and hybrid classifiers. International conference on pattern recognition (ICPR) 2008
2. Keysers D, Shafait F, Breuel TM (2007) Document image zone classification- a simple high-performance approach. In Proceedings of 2nd international conference on computer vision theory and applications 2007
3. Moll MA, Baird HS, Chang An (2008) Truthing for pixel-accurate segmentation. In: Document analysis systems. The eighth IAPR international workshop 2008
4. Shafait F, Keysers D, Breuel TM (2006) Pixel-accurate representation and evaluation of page segmentation in document images. In: 18th international conference on pattern recognition 2006
5. Bukhari SS, Ali AlAzawi M, Shafait F (2010) Document image segmentation using discriminative learning over connected components. In: 9th IAPR workshop on document analysis systems 2010
6. Bloomberg S, Chen FR (1996) Extraction of text-related features for condensing image documents. In: SPIE conference on 2660, Document Recognition III 1996
7. Bukharia SS, Shafaitb F, Thomas M (2011) Breuela: improved document image segmentation algorithm using multi-resolution morphology. SPIE Document Recognition and Retrieval XVIII 2011
8. Sarkar R, Moulik S, Das N, Basu S, Nasipuri M, Kundu M (2011) Suppression of non-text components in handwritten document images. International conference on image information processing (ICIIP) 2011
9. Otsu N (1979) A threshold selection method from gray-level histogram. IEEE Trans Syst Man Cybern
10. Ping Z, Lihui C, Alex KC (2000) Text document filters using morphological and geometrical features of characters. In: 5th international conference on Signal processing proceedings 2000
11. Di Stefano L, Bulgarelli A (1999) A simple and efficient connected components labeling algorithm. In: International conference on image analysis and processing ICIAP 1999
12. Granger E, Henniges P, Sabourin R, Oliveira LS (2007) Supervised learning of fuzzy ARTMAP neural networks through particle swarm optimization. J Pattern Recog Res
13. Taghi M, Baghmisheh V, Nikola P (2003) A fast simplified fuzzy ARTMAP network. J Neural Process Lett

A Novel Approach to Detect Anomalous Behaviour Using Gesture Recognition

Jeswanth Mohandoss

Abstract The primary goal of my work is to create a system which can identify specific human gestures and group the common gestures which in turn is used to convey information if uncommon activities are performed. Identification will be based on a video input based self learning gesture identification model which will classify the gestures based on genetic parameters. Most papers in this area focus on classifying different gestures, but do not judge whether the recognized gesture is good or bad in continuous recordings of daily life. The uniqueness of my approach lies in the method to manage a process of mass gesture detection in common places and classifying it using Support Vector Machine in the Learning mode. In the Execution mode, video footages are fed to my model which compares the current patterns with the stored normalized patterns and flag the ones that are odd.

Keywords Gesture recognition · Motion detection · Anomaly detection · Support vector machine

1 Introduction

A Gesture is a form of non verbal communication in which visible bodily actions communicate particular messages, either in place of speech or together and in parallel with spoken words. Gestures include movement of the hands, face or other parts of the body. Gestures differ from physical non verbal communication that does not communicate specific messages, such as purely expressive displays,

J. Mohandoss (✉)
Tata Consultancy Services, Karapakkam, Chennai, India
e-mail: jeswanthdoss@gmail.com

Mohan S. and S. S. Kumar (eds.), *Proceedings of the Fourth International Conference on Signal and Image Processing 2012 (ICSIP 2012)*, Lecture Notes in Electrical Engineering 222,
DOI: 10.1007/978-81-322-1000-9_11,

proxemics or displays of joint attention. Gestures allow individuals to communicate a variety of feelings and thoughts, from contempt and hostility to approval and affection, often together with body language in addition to words when the speak.

Gesture recognition is a topic in computer science and language technology with the goal of interpreting human gestures via mathematical works. Gestures can originate from any bodily motion or state but commonly originate from the face or hand. Current focuses in the field include emotion recognition from the face and hand gesture recognition. Many approaches have been made using cameras and computer vision to interpret sign language. However the identification and recognition of posture, gait, proxemics and human behaviours is also the subject of gesture recognition techniques.

Gesture detection can be seen as a way for computers to begin to understand human body language, this building a richer bridge between machines and humans than primitive text user interfaces or even GUI which still limit the majority of input to keyboard and mouse.

The main research work carried out in Image Processing is Face Recognition which in recent years has evolved as an uncomplicated task in this field. But applying it to a mass crowd to detect a gesture and recognize an uncommon activity or abnormality is quite a challenging task. In some gesture recognition systems, the states of an hidden Markov model represent visual units. A state transition defines the probability of the next state's occurrence. A complete specification of a hidden Markov model requires the state transition probability distribution, the observation symbol probability distribution, and the initial state distribution. Human tracking and gesture recognition has been an interesting topic in the research of human behavior understanding, human–machine interaction, machine control, surveillance, etc. In order to observe or analyze the characteristics of human motion and verify or evaluate the developed algorithm and its application system, the gesture database systematically constructed in carefully controlled environment is desperately required. Several gesture databases have been constructed for these purposes.

Crowded scenes are one such scenario of high-density, cluttered scenes that contain a large number of people. In such scenarios, finding every individual is a tough task and it depends on the quality of the video. Tracking objects or people is a crucial step in video analysis with a wide range of applications including behavior modeling and surveillance. Due to the large number of pedestrians in close proximity, crowded areas are at a high risk for dangerous activities including crowd panic, stampedes, and accidents involving a large number of individuals. The large amount of activity within such scenes is difficult for even human observers to analyze, making crowded scenes perhaps the most difficult task in automatic video analysis. An example crowded scene is shown in the Fig. 1.

Fig. 1 A crowded scene in which finding every individual's gesture is a tough task and it depends on the quality of the video

2 Related Work

There are several methods to track people in a scene. However the efficient method to track people in a crowded scene is by using Local Spatio-Temporal Motion [1]. For recognizing the actions performed by the objects, we can use Hierarchical Filtered Motion [2]. However since we are going to recognize Gestures, we have used Temporal Normalized Motion [3]. To detect pedestrian movement in specific, we might go for top-down segmentation [14]. Another approach in HSV Segmentation [4] is available to recognize gestures but it is complex and occupies more memory space than Temporal Normalized Motion. For recognizing hand gestures alone, there is a method in Bag of Features [5] but since we are going to detect full body gestures, it is used only for training using Support Vector Machine [5]. For tracking 'n' no of people in extremely crowded scenes, a better method is available in Dense Spatio-Temporal [6]. For human computer interaction, we have used a method of interpreting gestures visually [7]. Human movement can be distinguished from object movement using Temporal Templates [8]. If objects are hidden behind some static objects, we can recognize it using exemplar based Image Inpainting [9]. But a simple and novel replacement method for the previous approach is available with Temporal Segment Detection and Affect Recognition [10]. Gestures can be stored in Gesture database [11] but since we are going to consider velocity and frame rate per second, we use Behaviour matrices. Noise can be removed during preprocessing using 4th order Partial Differential Equations [13]. Anomalous behaviour may be detected using Bimodal Analysis [12] but it is easy to flag anomalies as odd ones.

3 Method

Our work has two modes, the Learning mode and the Execution mode. In both learning mode and execution mode, the input video is first preprocessed. The frames extracted after preprocessing is then compared with the reference image of a surveillance camera to recognize the objects and these images are stored in an array. The Preprocessed image array with recognized objects are then applied with Temporal Normalized Motion to recognize the gestures by observing changes after comparing a frame with the previous frame. Every Gesture of every person is recorded in Behaviour matrices along with the velocity and frame rate per second. These Behaviour matrices are stored in a Database during the learning mode. Every row of the Behaviour matrix is a pattern. In the execution mode, the current pattern of a person is matched with the normalized pattern stored in Behaviour Database and the odd ones are flagged with rectangular mark and timestamp. An Overview of our proposed system is explained in the Fig. 2.

3.1 Preprocessing

Human observers of crowds, particularly those experienced in the management of crowds in public places, can detect many crowd features, in some cases quite easily. Normally they can distinguish between a moving and a stationary crowd and estimate the majority direction and speed of movement of a large crowd, without needing to visually identify or count the separate individuals forming the crowd. They could also easily estimate in a qualitative way the crowd density to

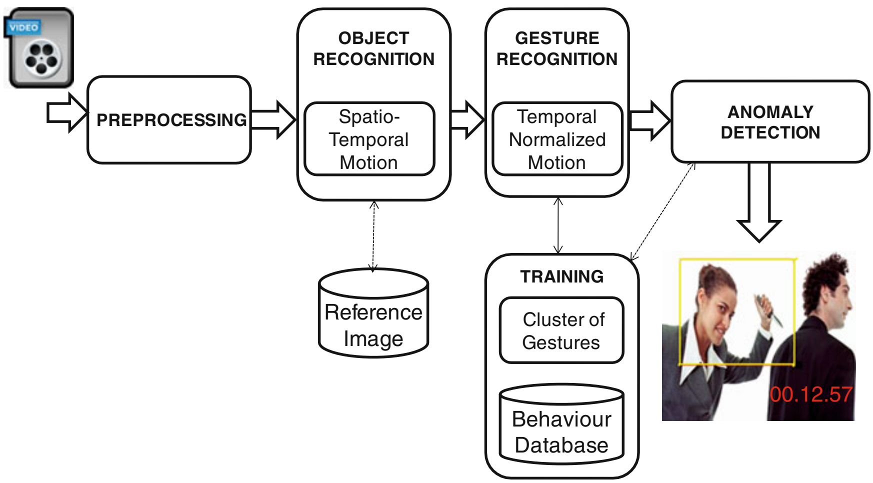


Fig. 2 Overview of the proposed approach

require image processing and computer vision techniques to match such capabilities of human observers is at present not unrealistic.

General Preprocessing steps include Extraction of Frames, Noise Removal, Garbage Value Removal and Leveling of Brightness for every frame so that it eases the process of Object Recognition. It is well-established that crowd congestion which is reaching a danger level can be spotted by observers noting that the up-and-down oscillatory head movements of individuals walking in a freely-flowing crowd stop when the crowd is too dense for free movement. Frequency-domain techniques are used to identify these up and down movements to discriminate between stationary and non-stationary flow.

A possible alternative is to compute the 2-dimensional discrete Fourier transform (DFT) for each image in a time sequence, followed by a measurement of temporal changes in the resulting magnitude and/or phase spectra (frequency domain). This approach has two main disadvantages. First, the DFT for a single image is related to local changes of intensity and not to temporal (inter frame) properties. Secondly, it involves a high computational and memory cost.

A more effective method is to isolate motion properties in the image sequence through a data-reducing coding mechanism. The preprocessing part is responsible for extracting frames from the input video and in adjusting the basic parameters. The input for Preprocessing is the video data and the output is the array of preprocessed frames. Noise refers to the randomized dot patterns that occur on the frame due to electronic sounds while extracting frames. 4th Order Partial Differential Equation (PDE) is used for Noise Removal by using the following equation.

$$u_t = -\Delta[c\,(\Delta u)\,\Delta u] \tag{1}$$

Preprocessing Block Diagram is represented in the Fig. 3. Preprocessing module functions using the following steps

- Extract one frame from the video feed.
- Remove noises and garbage values.
- Adjust the brightness throughout the image.
- Store it in an array.
- Fetch the next frame.

3.2 Object Recognition

The Object recognition part is responsible for identifying the objects in a frame. The input for Object Recognition is the array of preprocessed frames and the output are the frames with recognized objects. This unit uses Spatio-Temporal Motion to detect objects in a particular frame. Objects in the extracted frame are recognized by comparing it with the reference image. Motion detection is done by identifying pairs of pixel neighborhoods, in the two successive images, that have a

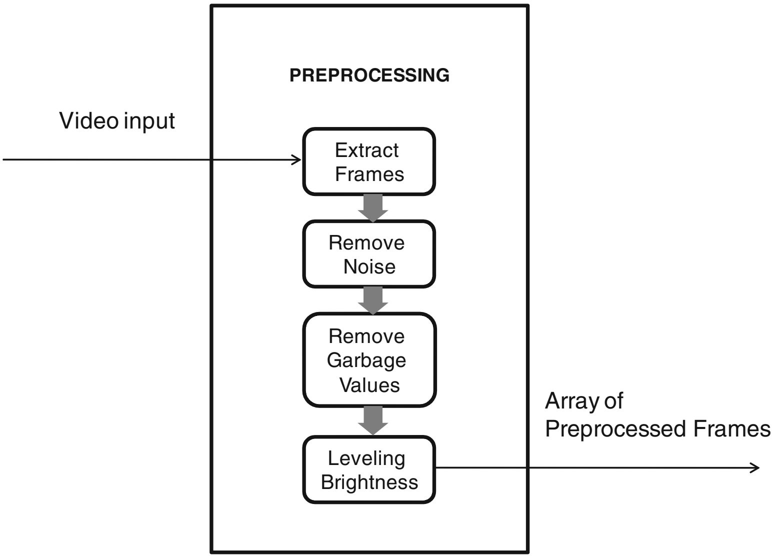


Fig. 3 Preprocessing block diagram

similar grey-level distribution. Pixels in the first image can, as before, be preselected to reduce the amount of data to process. This can be done using background removal (pixels in the background are not expected to move) or by computing the difference between the two images (a direct indication of motion). The objective is then to compute a velocity field for each preselected pixel (x, y) in the first image by defining a small neighborhood (or pixel block, typically 10×10) centered at position (x, y). A search block, also centered at (x, y) is defined in the second image. The size of the search block is determined by the maximum displacement expected in the given frame interval. Then, all possible pixel blocks in the search area are compared with that in the first image using a similarity function (e.g. sum of the pixel-to-pixel absolute difference). Under ideal conditions of no change in illumination and object shape, a matching block would be found where the similarity function is zero. Object Recognition Block Diagram is depicted in the Fig. 4. I defined the Spatio-Temporal Motion by 3D Gaussian distribution N (μ, Σ).

$$\mu = \frac{1}{N}\sum_{i}^{N} \Delta I_i \tag{2}$$

$$\mu = \frac{1}{N}\sum_{i}^{N} (\Delta I_i - \mu)\,(\Delta I_i - \mu)^T \tag{3}$$

where ΔI_i is the 3D Gradient Vector.

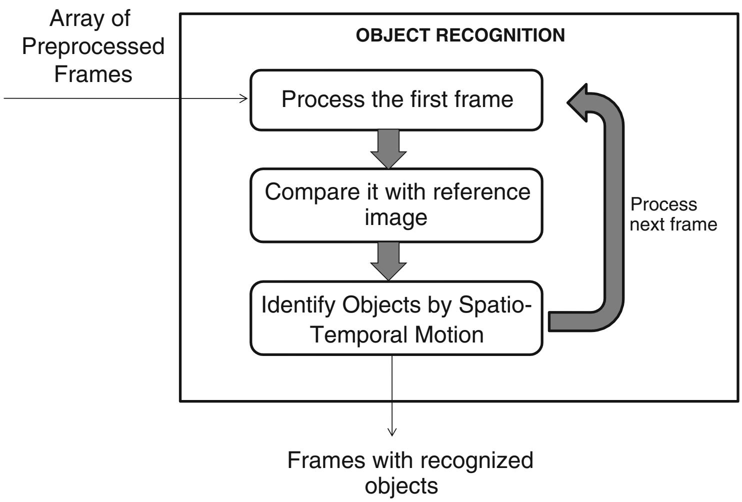


Fig. 4 Object recognition block diagram

Object Recognition works by using the following steps

- Get the preprocessed frame array input from the previous module.
- Process the first frame.
- Compare it with the Reference image.
- Recognize the objects using Spatio-Temporal Motion.
- Store the processed frames with recognized objects.
- Load the next frame.

3.3 Gesture Recognition

The Gesture Recognition is in charge for recognizing gestures of the identified persons. Gesture Recognition is done by Temporal normalized motion (TNM). The input for Gesture Recognition is the frames with recognized objects and the output are the behaviour matrices with recognized gestures. TNM method extracts the silhouettes from the frames with objects and uses the MHI-HOG and Image-HOG to recognize the face and body gestures respectively. MHI stands for Motion History Image and HOG stands for Histogram of Oriented Gradients. These HOGs are used in Automatic Storage Management (ASM) tracking to store the prototype. Then the HOG is passed to the HOG Descriptor which describes the HOGs as either MHI-HOG or Image-HOG. Then the HOGs are analyzed using Principal Component Analysis (PCA). This unit uses Temporal Normalized Motion to recognize gestures of a particular individual add it to its respective cluster using

Support Vector Machine (SVM) classifier and are stored in behaviour matrices along with its velocity v and frame rate per second fps. A set of safe and unsafe gestures are pre-defined.

Irregular motion, movements of arms, legs, and clothing and localized variations in brightness all cause errors in the computed motion vectors compared to the actual overall motion of the individuals in the crowd. To compensate for such effects (which can be effectively regarded as zero mean noise added to the motion vector estimates), the computed vectors are aggregated over small disjoint neighborhoods of 10×10 pixels throughout the image and gesture recognition is performed only when an anomaly from motion occurs. It also makes the complete process computationally more feasible. These recognized gestures of a person are then recorded in a Behaviour matrix B along with its velocity and frame rate per second (fps).

For example, if there are 10 people recognized in a frame, 10 rows are formed in the matrix with 3 parameters in columns. Every row having parameters is a pattern p. Gesture Recognition Block Diagram is represented in the Fig. 5. The Recognized Gesture G stored in a pattern p along with velocity v and frame rate per second fps is calculated by using the following formula

$$G = \Sigma + \mu\,\mu^{T} \tag{4}$$

The Gesture Recognition unit employs the following steps

- Get the frame array input with recognized objects from the previous module
- Load the first frame.

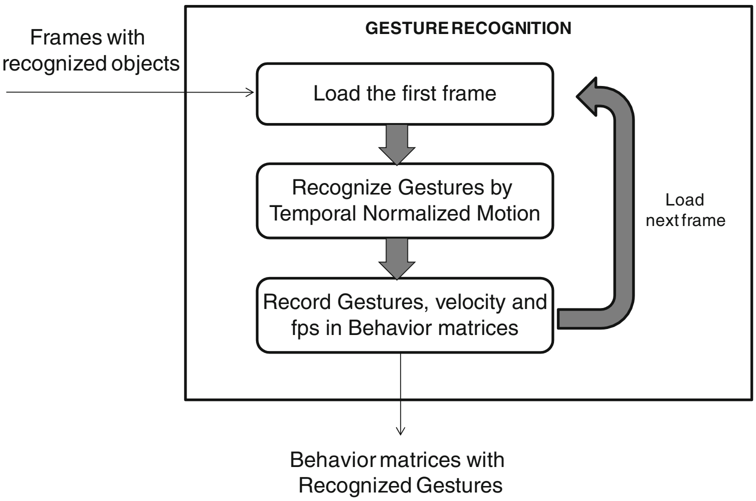


Fig. 5 Gesture recognition block diagram

- Apply Temporal Normalized Motion to identify Gestures by observing changes after comparing a frame with the previous frame and add it to the particular cluster.
- Record every Gesture of each person in Behavior matrix B along with velocity and fps.
- Process next frame in the array input.

3.4 Anomaly Detection

The Anomaly Detection unit is accountable for detecting any uncommon or suspicious behaviour. The input for this module is the Behaviour matrix B with recognized gestures and the output are the screenshots with anomalous behaviour. The Flowchart for Anomaly Detection Module is represented in the Fig. 6. The Anomaly Detection module makes use of the following steps

- Get the behavior matrices B with recognized gestures from the previous module
- Compare the current data pattern p of the input frame with the pool of stored normalized data patterns.
- If pattern match occurs,

(1) Check the next pattern in the matrix.

else

(2) Trigger by producing screenshots with Anomalous behaviour.

During execution, if an individual is identified in a frame, his pattern will be compared with the stored normalized patterns. Every gesture that is common will be left out. The odd ones are flagged and are displayed as anomalous behaviour in

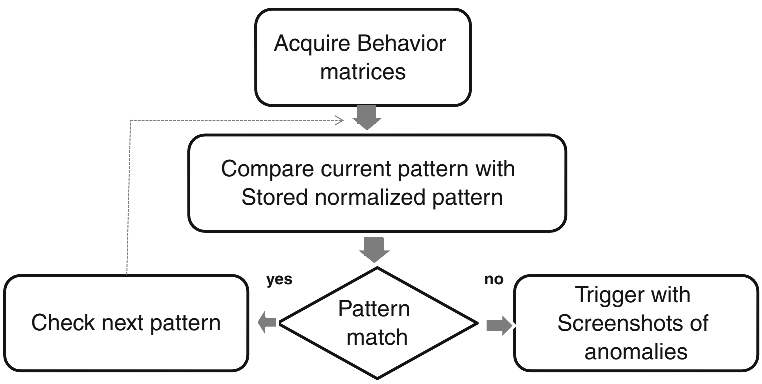


Fig. 6 Anomaly detection flowchart

screenshots with timestamp. I enumerate a function called Entropy that depends on motion area ratio, coefficient of direction variation, coefficient of distance variation, and direction histogram characteristics at any frame f. I examined and noted the similarities or differences of each calculated value of Entropy with a beforehand defined entropy threshold T_E, i.e., a deviant frame can be detected if and only if Entropy $< T_E$ otherwise standard frame.

The hypothetical outlook of reckoning T_E is that we pay attention to the minimum number of entropies in large videos that keep controls of standard events: n is the number of frames of the video database. The T_E (also Entropy) depends on the controlled environment (video stream), specifically the aloofness of the camera to the scene, the orientation of the camera, the type and the position of the camera, lighting system, density of the crowd (e.g., motion are a ratio, direction & distance variance-mean ratios, etc.), etc. The more is the remoteness between camera and the scene, the less is the considerable amount of optical flows and blobs.

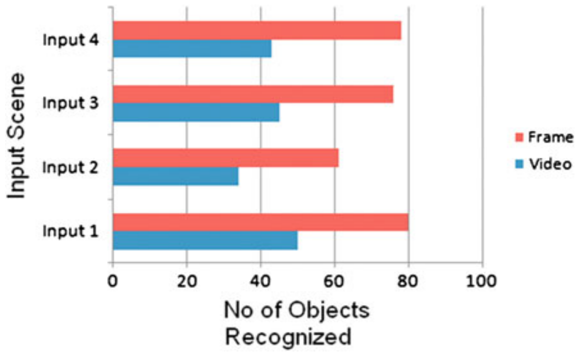
In case of escalator, T_E also places trust on the escalator type and position. Taking into account of these facts, we think carefully that we have at least one threshold by a video stream. If we have N video streams, then we put forward at least N thresholds. If the video stream leaves for another, then the threshold should be made over. The Formula for the Threshold T_E is given below.

$$T_E = \min_{k=1\ldots n}\{\text{Entropy}\}_k \tag{5}$$

4 Performance Analysis

The proposed system is directly given extracted frames of video file as input. Results showed that many irrelevant objects are recognized since the frames are not preprocessed. Instead if the direct video file itself is given as input to my system, most of the objects recognized are exact. Figure 7 shows the Objects recognized per frame. Grey represents video input and Black represents frame input.

Fig. 7 Video versus frame input



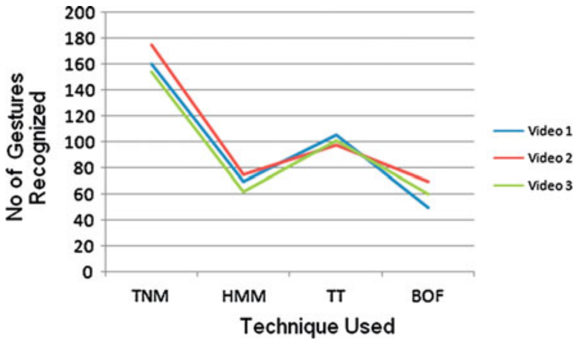


Fig. 8 Gestures recognized by SVM, HMM, TT and BOF

Figure 8 shows the Number of Gestures recognized per frame using SVM, HMM and TT for 3 unique Video inputs. My method of detecting Gestures of individuals recognized per frame is compared with other existing methods. It was inferred that Temporal Normalized Motion with Support Vector Machine (SVM) Training is able to detect almost every gesture per frame accurately than Hidden Markov Model (HMM), Bag of Features (BOF) and Temporal Templates (TT). Temporal Templates method was able to detect at most half of the gestures that are recognized by Temporal Normalized Motion. Bag of Features method was able to detect only hand gestures and Hidden Markov Models grouped many gestures under the same cluster.

4.1 Test Cases

There are 5 test cases that are given to our system. They are mentioned as follows

(Case 1) Video without any anomalous behaviour
(Case 2) Video with less crowded scenes
(Case 3) Video with extremely crowded scenes &
(Case 4) Video with single uncommon behaviour
(Case 5) Video with multiple uncommon behaviours

For the first test case, I got no screenshots since there is no anomalous behaviour to highlight. For the second test case, gestures of all the people are recognized clearly and Snapshots of uncommon behaviour if any are obtained. For the third test case, Gestures of all most all the people who are clearly visible are recognized and Snapshots of uncommon behaviour if any are obtained without any difficulties. For the fourth test case, we got a single snapshot of the uncommon behaviour as desired. For the final test case, multiple screenshots of various anomalous behaviours are obtained.

The snapshots of various anomalous behaviours found using our system is given in Figs. 9 and 10

Fig. 9 Anomaly- person is about to be hit by a car

Fig. 10 Anomaly- pedestrian walking in restricted area

5 Conclusion

I came up with a novel approach that inputs a video preprocess it and extract frames. Objects are identified in the extracted frames and its motion is detected. Gestures are then recognized using Temporal Normalized Motion and these patterns are recorded in Behaviour matrices in the Learning mode. In the execution mode, current patterns are matched with stored patterns and the odd ones are displayed as anomalous behaviour. This approach with some changes according to the scene in which the camera is installed can be used in surveillance systems. Rather than wasting time in watching whole videos recorded by the surveillance camera, if we input the recorded video files to our system, it simply produces screenshots of anomalous behaviour along with timestamp. The accuracy level of the proposed system is much more than any other existing systems.

Acknowledgments The author is grateful to Dr. S. Chitrakala M.E., Ph.D., for motivating me to present my work in conferences. I would like to thank my mentor Mr. M. Arun Marx M.E., for standing with me through my tough times and supporting me constantly.

References

1. Kratz L, Nishino K (2012) Tracking pedestrians using local spatio-temporal motion patterns in extremely crowded scenes. IEEE Trans Pattern Anal Mach Intell 34(5):987–1002
2. Tian YL, Cao L, Liu Z, Zhang Z (2012) Hierarchical filtered motion for action recognition in crowded videos. IEEE Trans Syst Man Cybern Part C Appl Rev
3. Chen S, Tian YL, Liu Q (2011) Recognizing face and body gesture by temporal normalized motion and appearance features. IEEE computer society conference on computer vision and pattern recognition workshops, June 2011
4. Hasan MM, Misra PK (2011) Gesture recognition using modified HSV segmentation. In: International conference on communication systems and network technologies (CSNT), June 2011
5. Dardas NH, Georganas ND (2011) Real-time hand gesture detection and recognition using bag-of-features and support vector machine techniques. IEEE Trans Instrum Meas 60(11):3592–3607
6. Hofmann M, Rigoll G, Huang TS (2010) Dense spatio-temporal motion segmentation for tracking multiple self-occluding people. IEEE computer society conference on computer vision and pattern recognition workshops, June 2010
7. Pavlovic VI, Sharma R, Huang TS (1997) Visual interpretation of hand gestures for human-computer interaction: a review. IEEE Trans Pattern Anal Mach Intell 19(7):677–695
8. Bobick AF, Davis JW (2001) The recognition of human movement using temporal templates. IEEE Trans Pattern Anal Mach Intell 23(3):257–267
9. Criminisi A, Perez P, Toyama K (2004) Region filling and object removal by exemplar-based image inpainting. IEEE Trans Image Process 13(9):1200–1212
10. Gunes H, Piccardi M (2009) Automatic temporal segment detection and affect recognition from face and body display. IEEE Trans Syst Man Cybern Part B Cybern 39(1)
11. Hwang B-W, Kim S, Lee S-W (2006) A full-body gesture database for automatic gesture recognition. In: International conference on automatic face and gesture recognition, April 2006

12. Gunes H, Piccardi M (2006) A bimodal face and body gesture database for automatic analysis of human nonverbal affective behavior. International Conference on Pattern Recognition, 2006
13. Lysaker M, Lundervold A, Tai X-C (2003) Noise removal using fourth-order partial differential equation with applications to medical magnetic resonance images in space and time. IEEE Trans Image Process 12(12):1579–1590
14. Leibe B, Seemann E, Schiele B (2005) Pedestrian detection in crowded scenes. IEEE computer society conference on computer vision and pattern recognition, vol 1, June 2005

ECG Beats Extraction and Classification Using Radial Basis Function Neural Networks

Mohammed Belkheiri, Zineb Douidi and Ahmed Belkheiri

Abstract This paper aims the design of an ECG diagnosis system that helps physicians in the interpretation of ECG signals. This system preprocesses and extracts the ECG beats of an ECG record and some feature extraction techniques are invoked to get a feature vector that represents the main characteristics of the ECG wave. After that a well trained RBF artificial neural network is used as a classifier for four different ECG heart conditions selected from MIT-BIH arrhythmia database. The ECG samples were processed and normalized to produce a set of reduced feature vectors. The results of sensitivity, specificity accuracy and recognition rate of the system are analyzed to find the best RBF neural network for ECG classification. Different ECG feature vectors composed of averaged amplitude values, DCT coefficients, DFT coefficients, and wavelet coefficients were used as inputs to the neural network. Among different feature sets, it was found that an RBF which has one layer and the feature vector 61 inputs, and 20 neurons possessed the best performance with highest recognition rate of 95 % for four cardiac conditions.

Keywords ECG · Classification · DWT · DCT · DFT · RBF neural networks · Feature extraction

M. Belkheiri (✉) · Z. Douidi · A. Belkheiri
Laboratoire de Télécommunications, Signaux et Systèmes, Department of Electronics, Faculty of Technologie, University Amar Telidji of Laghouat, BP G37 Route de Gardaia 03000 Laghouat, Algeria
e-mail: m.belkheiri@mail.lagh-univ.dz

Mohan S. and S. S. Kumar (eds.), *Proceedings of the Fourth International Conference on Signal and Image Processing 2012 (ICSIP 2012)*, Lecture Notes in Electrical Engineering 222,
DOI: 10.1007/978-81-322-1000-9_12,

1 Introduction

Electrocardiogram is an important tool for providing useful information about the heart activity and it is widely used as a monitoring tool that helps physicians during their diagnosis. Today thanks to the advances in electronics and computing that makes a variety of devices available for ECG signals acquisition and processing to help in the interpretation of the information behind the ECG graph.

Many researchers in applied signal processing try to design automated systems based on linear and nonlinear filtering, pattern recognition and artificial intelligent systems that can extract the useful information in such signals and provide it as a digital health monitoring system to help in saving the human life.

The schematic of a single heartbeat in ECG signal is indicated in Fig. 1. The normal kind of signal belonged to a healthy person is according to a known structure, changing and disturbing in any important parameters represent a heart disease.

The ECG waveform has various components such as waves, segments, and intervals that are evaluated and classified based on their size, length of time, and location on the tracing. All of these different components determine the type of cardiac rhythm.

The recorded graph of the ECG waveform produced by the heart can tell us basic information about a patient's condition. The ability to evaluate various ECG waveforms is an important skill for many health care professionals including nurses, doctors, and medical assistants.

Many papers appeared in the literature presenting a variety of signal processing algorithms developed for the recognition and classification of ECG waves. These techniques either use temporal or frequency domain analysis for feature extraction allowing the recognition between the beats belonging to different pathological classes.

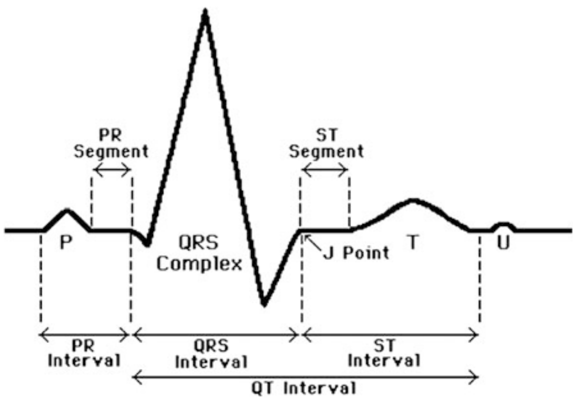


Fig. 1 Schematic of ECG signal

Artificial Neural Networks (ANN) and fuzzy-based techniques were also employed to exploit their natural ability in pattern recognition task for successful classification of ECG beats [1, 2].

Many artificial intelligent techniques such as maximum likelihood, artificial neural networks [2], and support vector machines [3] have been used successfully for ECG beat classification. These methods based on supervised learning learn from the samples of training data and map new data instances based on the information extracted from the annotated training data (targets) [4].

The wavelet transformation provides another alternative for researchers in bioinformatics to extract valuable information from ECG signal based on extracting and representing the signal in different levels and translations in both frequency and time domain. The extracted parameters from the transformed coefficients are capable to represent in a relevant manner the main characteristics of the original signal [5].

Among the tested classifiers like support vector machines (SVM), artificial neural networks and fuzzy logic techniques, researchers reported that the techniques based on neural networks represent a powerful tool that has been applied successfully for ECG beat classification for the remarkable properties that it provides like adaptability, learning from examples and generalization [4, 5].

In this paper, ECG beat classification is conducted on three steps:

- Detecting, normalizing and centering one beat in a window of 256 samples from a record of several beats,
- Features extraction based on linear and non linear filtering to reduce the input space of the classifier,
- Classifier design using RBF neural Networks.

Further, the performance of the RBF Neural Network classification accuracy is evaluated for generalization.

The rest of the paper is organized as follows. The method for ECG feature extraction, the basic mathematical formulation of RBF for solving classification problems, and the working methodology are given in Sects. 2 and 3. The classification results obtained on ECG data from the standard (MIT-BIH) arrhythmia database [6] are presented and discussed in Sect. 4. Finally, conclusions and future work are reported in Sect. 5.

2 Methods

In this study we have chosen Radial Basis Function (RBF) neural networks for ECG waves classification for a data set that consists of 1,280 vectors taken from 8 persons that 25 % of them were healthy and 75 % of them were patient. As illustrated in Fig. 2, there are three major steps to the ECG signal pattern recognition, namely, Pre-Processing of the signal, ECG feature extraction and ECG signal classification.

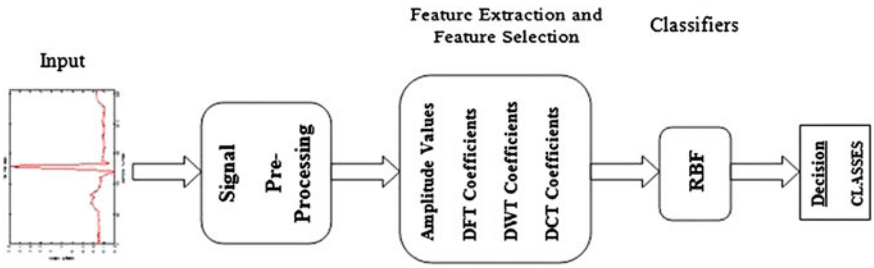


Fig. 2 Block diagram of ECG classification

The first step is the measurement of acquisition period, which requires a wide range of the ECG signal collection including different abnormalities. The data is available from the MIT-BIH Arrhythmia database [6].

The second step is to find the smallest set of features that maximize the classification performance of the next step. ECG feature extraction is mainly used in this step. The artificial neural network will be used in this paper to do the ECG classification.

In feature extraction we used two techniques to extract the features: direct technique which include ECG amplitude values and transformation technique which include three transformation techniques Discrete Fourier Transform (DFT), Discrete Cosine Transform (DCT) and Discrete Wavelet Transform (DWT). In the feature selection we used the results of the paper [3].

3 Pre-Processing and ECG Beat Detection

ECG signals are quasi-periodic of relatively low amplitude of several mV. They are often affected by noise. The acquisition of the signal requires their amplification and further processing in order to suppress noise to highest extent [2, 7]. Further ECG signal analysis is realized based on these recordings for which the noise due to power-line interferences, EMG artifacts from muscles, motion artifacts from the subject and electrodes has been suppressed.

In the sequel the digital ECG signal is filtered to eliminate noise and further processed to enhance effectiveness of methods of feature selection, classification, and interpretation applied to the signal.

After retrieving signals from MIT-BIH Arrhythmia database the ECG signals are filtered with band pass filter of (1–100 Hz) [7]. In order to detect and isolate one heart beat a simple algorithm based on R-wave peak detection process is employed on ECG data as suggested in [3].

A window of 256 data samples is formed by centering the R peak in the QRS complex for a single ECG beat pattern. Peak-to-Peak amplitude of the ECG signal are normalized to a value of 1 mV. Thus, it is noted that classification decision does not depend on the maximum amplitude of the ECG records. Mean value of ECG signal in the window is fixed to zero value by subtracting mean value from original signal. Thus, offset effect is canceled out.

3.1 Feature Extraction

Automatic ECG beat recognition and classification are performed relying on various features, time domain representation, extracted from the ECG beat, or the measure of energy in a band of frequencies in the spectrum (frequency domain representation).

In this contribution, following the same strategy used in the paper [3] for feature extraction using divergence analysis where four different classes of feature sets are combined to construct the feature vector belonging to the isolated ECG beats including averaged temporal values, Direct Cosine Transform coefficients, DFT Fourier coefficients and discrete wavelet transform detail coefficients for the different scales (1–4 scales).

The first feature set is composed of averaged amplitude values of the 256 samples of one ECG beat where the number of samples is reduced simply to 16 as follows:

$$k_j = \frac{1}{16}\sum_{i=1}^{16} x\,(i + (j-1)\,16), \quad j = 1, 2, \ldots, 16 \tag{1}$$

where $x(1)$, $x(2)$,...,$x(256)$ are the samples of the original ECG pattern, and k_j is the jth element of the new feature vector formed by taking the average of every 16 samples of the original ECG signal.

The second feature extraction method is based on taking the 15 most discriminate coefficients of the discrete Fourier transform (DFT) of a window of 256 data length around the centred R-peak.

The frequency-based elements of each feature vector are ordered as follows:

$$K_2 = [f_1\,f_2\,f_5\,f_6\,f_8\,f_{11}\,f_{12}\,f_{13}\,f_{16}\,f_{20}\,f_{33}\,f_{35}\,f_{38}\,f_{40}\,f_{45}] \tag{2}$$

K_2 is the 15-dimensional new feature vector containing only 15 DFT coefficients from the vector F representing all the DFT coefficients

$$F = [f_1 f_2 \ldots f_{256}] \tag{3}$$

The third feature set involves the discrete cosine transform (DCT) coefficients of ECG data which are chosen from calculated DCT coefficients C, the 15 DCT coefficients are selected.

$$C = [c_1\,c_2 \ldots c_{256}] \tag{4}$$

The chosen elements in terms of significance level for DCT coefficients are arranged in K_3 as follows:

$$K_3 = [c_1\ c_2\ c_4\ c_{10}\ c_{11}\ c_{13}\ c_{17}\ c_{23}\ c_{25}\ c_{28}\ c_{29}\ c_{30}\ c_{32}\ c_{33}\ c_{35}] \tag{5}$$

where c_i is the ith element of the C DCT vector.

The feature extraction of fourth feature set involves the DWT of ECG data. Feature vectors are formed using Daubshies-2 wavelet [7]. For each feature vector, wavelet approximation coefficients at the second, third and fourth levels (66 + 34 + 18 = 118) are calculated and arranged in vector A as follows:

$$A = \left[a_2^1, \ldots a_2^{66},\ a_3^1, \ldots a_3^{34}, a_4^1, \ldots a_4^{18}\right] \tag{6}$$

where A is a vector representing the DWT approximation coefficients at second, third and fourth levels. Among the 118 elements of DWT approximation coefficients, a subset of 15 DWT coefficients are chosen to construct 15-dimensional new feature vector:

$$K_4 = \left[a_4^1\, a_4^2\, a_4^3\, a_4^4\, a_4^5\, a_4^6\, a_4^7\, a_4^8\, a_4^9\, a_4^{10} a_4^{11}\, a_4^{12}\, a_4^{13}\, a_4^{14}\, a_4^{15}\right] \tag{7}$$

Finally a set of 61 combined feature elements are combined to construct one feature vector K for each heart beat as follows:

$$K = [\,K_1 \quad K_2 \quad K_3 \quad K_4\,] \tag{8}$$

K is constructed to form the feature vector that will be used in classification which is different from the strategy of the paper [3] where the authors used each feature set K_i to design a separated classifier.

3.2 Classification Using RBF Neural Networks

Thanks to the universal approximation property of neural networks of nonlinear mappings between the input space and the output space. Artificial neural networks are applied successfully in several fields such as medicine, physics, electrical engineering, etc. to get simulation models for complex phenomena where physical laws can not give satisfactory models [8].

In this paper, RBF neural networks are used to solve the classification problem of ECG signals. Figure 3 depicts a standard RBF Neural Network. Assume that there exists an artificial neural network as shown in Fig. 3 that works as a classifier that maps the input feature space $\mu = [k_1, \ldots, k_{61}]$ to the class space $C_d = [\,C_1 \quad C_2 \quad C_3 \quad C_4\,]$.

The classifier output can be expressed as:

$$C_d \;=\; W^T\,\Phi\,(V, \mu) \;+\; \varepsilon(\mu), \forall \mu \in D \tag{9}$$

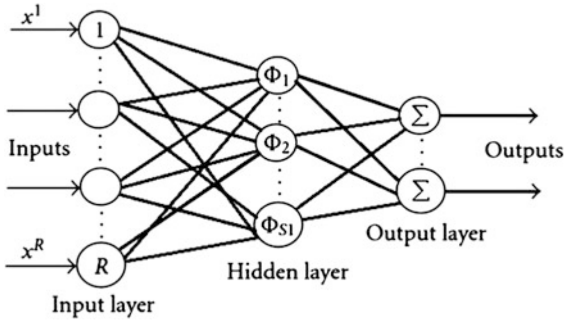


Fig. 3 An RBF neural network

where

μ is the input vector,
$V \in D_V \subset \Re^{N_1}$ is the hidden layer weight vector,
$W \in D_W \subset \Re^{N_2}$ is the output layer weight vector,
Φ is a set of Gaussian basis functions, and
ε is the Neural Network reconstruction error.

An RBFNN is composed of two layers; a hidden layer contains a set of neurons with their associated centers, the output of each neuron gives the distance between the input vector μ and its center μ_i. The output of the RBF network is a linear combination of the outputs of the hidden layer. It is given by:

$$\psi(\mu_j) = \sum_{i=1}^{N_2} w_i \varphi_{ij} \tag{10}$$

where $\varphi_{ij} = \exp\left(\frac{\mu_i - \mu_j}{\sigma}\right)$ is a Gaussian activation function of the distance between the input μ_i and the jth center μ_j.

The centres and widths of RBFNN are the two parameters which can affect the classification performance. Several methods have been proposed to find the centres of the RBFNN. These are usually clustering based methods that find centre locations among the input feature vector locations by minimization of some criteria or a subset of the input feature vectors can be picked directly from the training data base and used as the centers of the neurons. Hence it has been confirmed that the best center locations may not be necessarily located inside the input feature vectors. The most common algorithm to determine the neuron centers of the RBFNN are the K-Means algorithm.

So the objective is to construct and train an RBF neural network that can read the input feature vector corresponding to a given heart beat class and gives the binary-coded desired values of (0,0,0,1), (0,0,1,0), (0,1,0,0), (1,0,0,0) that corresponds to the four classes normal beat (N), atrial fibrillation beat (AF), paced beat (P) and sinus arrhythmia beat (SA) respectively as illustrated in Fig. 4.

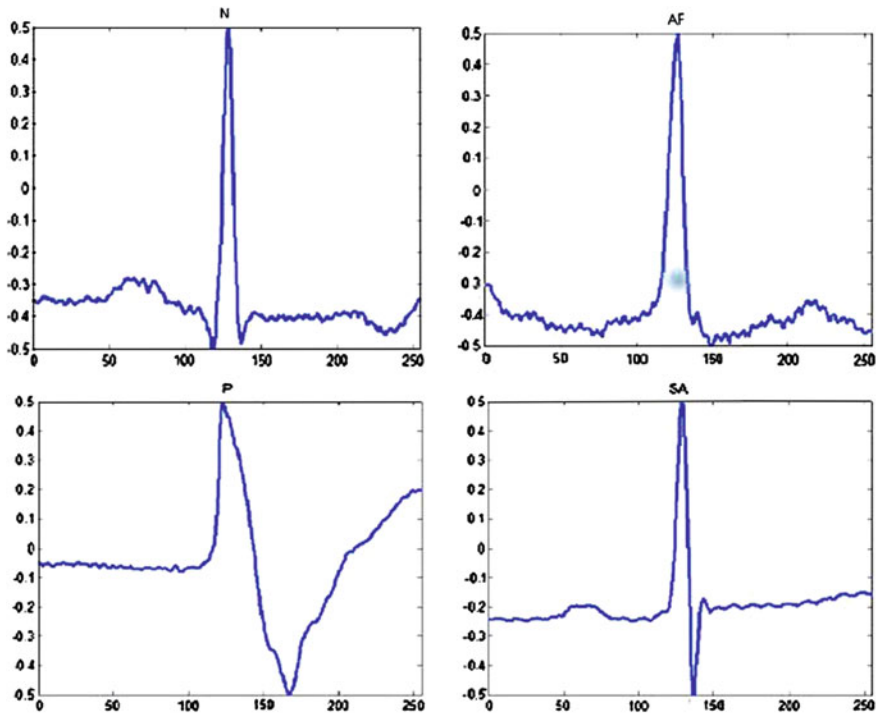


Fig. 4 The four different classes of ECG beats

4 Results and Discussions

The training and validation data set is formed by choosing 1,280 vectors (320 vectors for each class) of 256 dimensions representing four (04) classes from the standard MIT-BIH database [6]. Six hundred and forty vectors (160 from each class) are used for testing procedure. Six hundred and forty vectors (160 from each class) for testing procedure. In addition, in order to have a better generalization ability of the proposed system, the training and testing sets are formed from different subject's records.

The system is evaluated using one channel and one-dimensional 1,280 vectors (320 vectors from each class). 640 vectors (160 vectors from each class) of them are used for training procedures. The remaining 640 vectors (160 for each class) are used for testing procedures. The classified beats are normal beat (N), atrial fibrillation beat (AF), paced beat (P) and sinus arrhythmia beat (SA).

The output of the classification using an RBF classifier after training is highlighted in Fig. 5 for four different ECG beats showing that the classifier is performing well.

Table 1 shows the performance achieved using the trained RBF classifier in terms of recognition rate.

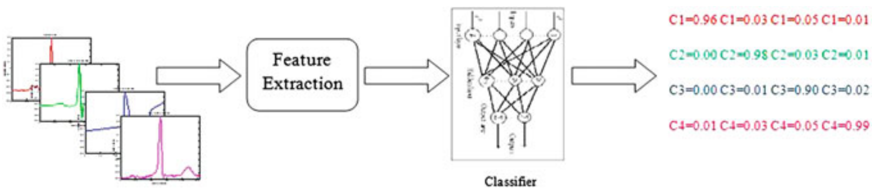


Fig. 5 The four different ECG beats correctly classified by the RBF Neural Network

Table 1 Recognition rates of RBF neural network classifier for different classes

ECG Class	Recognition rate (%)
C_1	97.2
C_2	91.7
C_3	99.5
C_4	93.1

Table 2 Sensitivity, Specificity and Accuracy for RBF

ECG Class	Sens. (%)	Spec. (%)	Accur. (%)
C_1	93.9	89.2	91.7
C_2	96.4	82.3	93.4
C_3	91.5	88.8	89.8
C_4	96.7	92.6	93.6
Average	94.6	88.2	92.0

This result shows that the recognition rate is more than 90 % for all the data classes. It is very acceptable for automatic classification applications which can be accepted in the medicine diagnostic area. This means that the training result was satisfactory and caused the testing result to be acceptable.

The performance of the system for the validation data set composed of records of patients that are not used during training in terms of sensitivity, specificity and accuracy for RBF classifier; they are shown in Table 2.

5 Conclusion

In this contribution, we have presented an automatic system for ECG beat classification that may be used as a diagnosis tool in medicine. This system detects a window of 256 samples of one heart beat then extracts its main temporal, frequency, and DWT features. These features are further processed by an RBFNN classifier. This system is designed and trained using records from the MIT-BIH Arrhythmia database to differentiate between four different heart conditions.

The advantage of the RBFNN classifier using the proposed feature extraction method is its simplicity, ease of implementation and its generalization ability. The best highest achieved performance of the RBFNN classifier is a recognition rate of 95.38 % for the newly introduced data which is comparable with many resent studies in this field. This contribution may be extended to classify other heart conditions.

References

1. Ge D, Xiao Q (2006) Feature extraction based on optimal discrimination plane in ECG signal classification. In: ADMA 2006.Lecture notes on artificial intelligence, vol 4093, pp. 143–149
2. Olmez T, Dokur Z (2003) Application of InP neural network to ECG beat classification. Neural Comput Appl 11:144–155
3. Acir N (2005) Classification of ECG beats by using a fast least square support vector machines with a dynamic programming feature selection algorithm. Neural Comput Appl 14(4):299–309
4. Faezipour M, Saeed A, Nourani M (2010) Automated ECG profiling and beat classification. In: IEEE international conference on acoustics speech and signal processing, pp 2198–2201
5. Hendel M, Benyettou A, Hendel F, Khelil H (2010) Automatic heartbeats classification based on discrete wavelet transform and on a fusion of probabilistic neural networks. J Appl Sci 15:1554–1562
6. MIT-BIH Arrhythmia Database, Physiobank Archieve. http://www.physionet.org/physiobank/database
7. Korurek M, Dogan B (2010) ECG beat classification using particle swarm optimization and radial basis function neural network. Expert Syst Appl 33:7563–7569
8. Haykin S (1998) Neural networks: a comprehensive foundation. Prentice Hall, New Jersey
9. Karpagachelvi S, Arthanari M, Sivakumar M (2011) Classification of electrocardiogram signals with support vector machines and extreme learning machine. Neural Comput Appl 20:1043–1053

Gender Classification Using Ear Biometrics

P. Gnanasivam and S. Muttan

Abstract In this work, ear biometrics has been used for gender classification. Identifying a person as male or female is an interesting problem and is required in many practical applications. The earhole has been considered as the primary reference point. Relative distances (Euclidean distance) have been measured between the ear identification points (ear features) and the ear hole. The ear features considered are outer lobe edge, outer and inner curves of the helix, outer and inner curves of the antihelix and two edges of the concha. We have used an extensive internal database of about 342 samples of male and female ears. The Bayes classifier, K-Nearest Neighbour (KNN) classifier and the neural network classifier have been used for the classification. Overall classification rate of 90.42 % is achieved using KNN classifier.

Keywords Biometric · KNN · Bayes classifier · Gender classification

1 Introduction

Gender detection is a very useful task for a wide range of applications. Gender classification has been especially interesting for psychologists. But a robust gender classification system could provide basis for performing passive surveillance using

P. Gnanasivam (✉)
Department of ECE, Agni College of Technology, Thalambur, OMR, Chennai, India
e-mail: pgsivam@gmail.com

S. Muttan
Centre for Medical Electronics, Department of ECE, Anna University,
Guindy, Chennai, India
e-mail: muthan_s@annauniv.edu

Mohan S. and S. S. Kumar (eds.), *Proceedings of the Fourth International Conference on Signal and Image Processing 2012 (ICSIP 2012)*, Lecture Notes in Electrical Engineering 222,
DOI: 10.1007/978-81-322-1000-9_13,

demographic information. Automatic gender classification also has applications such as human identification, smart human–computer interface, computer vision, passive demographic data collection, etc. The gender detection is useful in audio segmentation system for speaker clustering.

The ear is considered as an alternative to be used separately or in combination with other biometric entities as it is comparatively less affected by factors such as mood, health, and clothing. Significantly, the appearance of the auricle (outer ear) is relatively unaffected by aging, making it better suited for long-term identification when compared to other non-invasive techniques such as face recognition. Ears are a promising avenue for biometrics research and have been found to be able to identify subjects with similar performance to face recognition [1]. It is proved [2] that the ear is fully formed at birth. Even though the lobe descends a little, but overall it stays the same. Ear is more consistent compared to face as far as variability due to expression, orientation of the face and effect of aging [3]. The research done in Southampton University in November 2010 establishes ear biometrics as a better alternative to face and fingerprint recognition [4].

In this work, the earhole is considered as the primary reference point. The seven points of identification selected for the gender classification are outer lobe edge, outer and inner curves of the helix, outer and inner curves of the antihelix and two edges of the concha. KNN, Bayes and NN are used as classifiers. A block diagram of the gender classification method proposed is shown in Fig. 1.

2 Past Work in Ear Biometrics

Gender classification using ear biometrics has not been tried and most of the works are in the personal identification. Few research papers related to the personal identification is discussed in the section. Many researches on automatic detection of ear from the side face as enrolment is carried out for the personal identification [5–9]. One of the first ear recognition systems is the Iannarelli's system which was originally developed in 1949 [10]. There is existing surveys on ear describing feature extraction methods for ear biometric systems [11–14]. Victor et al. [15] and Chang et al. [16] have used principal component analysis (PCA) and FERET evaluation protocol for their research about the ears. Burge and Burger [17, 18] presented one of the most cited ear biometric methods in the literature.

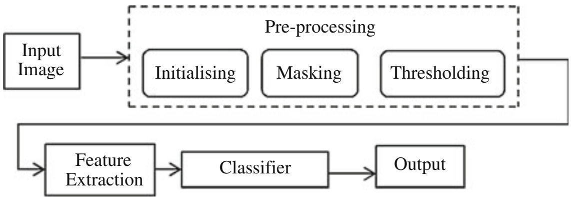


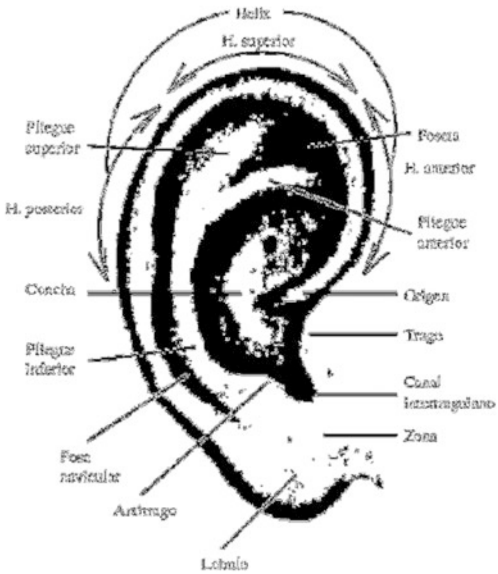
Fig. 1 Block diagram of gender classification method

Hurley et al. [19] have used force field transformations for ear recognition. Ansari and Gupta [20] used outer helix curves of ears moving parallel to each other as feature for localizing ear in an image. Haj Said et al. [21] addressed the problem of a fully automated ear segmentation scheme by employing morphological operators. Prakash et al. used skin-color and template based technique for automatic ear detection in a side profile face image [22, 23]. Yan and Bowyer [24] conducted a small experiment to test the ear symmetry using 119 subjects and Abaza and Ross [25], presented a detailed analysis of the bilateral symmetry of human ears. Related to the Forensic science, the ear marks left by the suspect, when presses their ear against a wall, window or door is found close to 15 % of crime scenes [26]. There have been several court cases in the US and other countries where earprints have been used as physical evidence [27, 28].

3 Ear Anatomies and Pre-processing

The human ear or the pinna is shown in Fig. 2. In addition to the conspicuous rim or the helix, the ear also has other prominent features such as the anti-helix which runs parallel to the helix, and a distinctive hairpin-bend shape just above the lobe called the inter notch. The central area or concha is named for its shell-like appearance.

Fig. 2 Ear anatomy



3.1 Pre-processing

Images with ear rings, other artefacts and occluded with hairs have not been processed in this research work. Each image is gone through the following steps before feature extraction. Figure 3 shows the output of the pre-processing steps.

- Ear image is cropped manually and resized.
- Colored image is converted to grayscale image.
- Adaptive thresholding and masking.

3.2 Thresholding

Thresholding may be viewed as an operation that involves tests against a function of the form,

$$T = T[(x, y, p(x, y)f(x, y)] \tag{1}$$

Where, $f(x, y)$ is the grey level of point (x, y) and $p(x, y)$ denotes some local property of this point. A threshold image $g(x, y)$ is defined as,

$$G(x, y) = \begin{cases} 1 \text{ } if \text{ } f(x, y) > T \\ 0 \text{ } if \text{ } f(x, y) \leq T \end{cases} \tag{2}$$

When T depends on $f(x, y)$, the threshold values are called global. If T depends on $f(x, y)$ and $p(x, y)$, the threshold is called local.

3.2.1 Adaptive Thresholding

First, the image histogram is partitioned using a single global threshold, T. Segmentation is then accomplished by scanning the image pixel by pixel and labelling each pixel as object or background, depending on whether the grey level of that

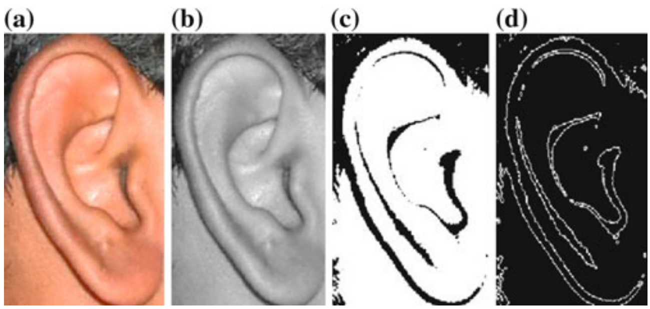


Fig. 3 Pre-processing output of ear image **a** Original image **b** Grayscale image **c** Thresholded image **d** Masked image

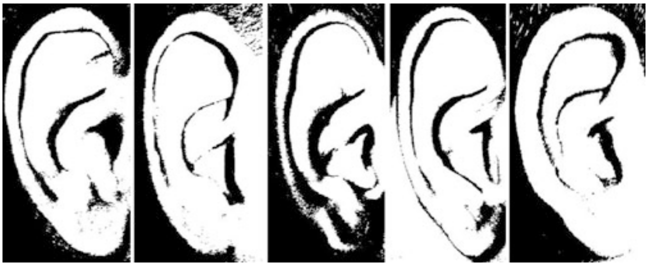


Fig. 4 Sample thresholded images

pixel is greater or less than the value of T. An algorithm for calculating T automatically is given as:

- Segment the image with initial estimate of T. This will produce two groups of pixels: G_1 consisting of all pixels with gray level values $> T$ and G_2 consisting of pixels with values $\leq T$.
- Compute the mean values μ_1 and μ_2 for the pixels in regions G_1 and G_2.
- Compute a new threshold value:

$$T = 0.5(\mu_1 + \mu_2) \tag{3}$$

- Repeat above steps until the difference in T in successive iterations is smaller than a predefined parameter T_0.

Adaptive thresholding is done by dividing the original image into two sub images and then utilizing a different threshold to segment each sub image. This kind of thresholding can accommodate changing lighting conditions in the image, e.g., those occurring as a result of a strong illumination gradient or shadows. Figure 4 shows the samples of thresholded images.

After the threshold is determined by a series of iterations, gray level values at each pixel are rounded up or down according to the equation:

$$F(x, y) = \begin{cases} 255 \text{ (white) } \mathit{if}\, f(x, y) \geq T \\ 0\,(\mathit{black})\ \mathit{if}\ f(x, y) < T \end{cases} \tag{4}$$

4 Ear Feature Extractions

4.1 Locating the Center of the Ear Hole

The ear canal (its outer center is referred as ear hole) is a tube running from the outer ear to the middle ear. The canal is approximately 5–10 mm in diameter and is approximately center to the middle line of the ear image. The ear hole is considered as a reference point to find Euclidean distances between all features.

- After the image is cropped, the closed loops of the image are filled to produce continuous white regions, one of which will be the earhole.
- Traversing from right to left, the first occurrence of a white pixel is noted.
- By iterative curve tracing processes, the top and bottom points of the earhole are identified.
- The mean of the ordinates gives the horizontal bisector of the hole, and the mean of the abscissas of the points of transition will be the center of the earhole, and the reference point for our consideration.

4.2 Seven Features of Ear

A morphological structuring element is created to construct a comprehensive image. The discontinuities in the created image are removed by connecting the disconnected points. The light structures attached to the borders are also suppressed. Figure 5 illustrates the extracted features.

The size of the image, i.e., the number of rows and columns in the image considered is determined. The values for the horizontal and vertical extremities and various other points are initialized. The horizontal left extreme is determined by checking each and every point for a pixel value of 255. After a point is determined, it is ensured that no more edges are found towards its left. Similarly the horizontal right, vertical top and bottom points are also determined. Similarly by locating points toward the right of the midpoint of the horizontal length, the hole of the ear is determined. From the right end of the ear hole (right side ear), the next occurrence of the white pixel is noted which is the point in antitragus. Similarity the various points are identified by iterative tracing the image in horizontal and vertical direction for the occurrence of white pixel. The point is

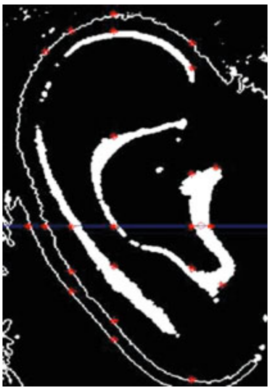


Fig. 5 Image showing identification of standard features

confirmed by locating the top and bottom points and ensuring that the height lies within the desired values. From all the calculated values, the anti-helix, concha, left end, the centre, antitragus, helix bottom, helix top, helix lobe are determined.

5 Classifiers Used in Gender Tagging

All classification algorithms typically employ two phases of processing: training and testing. In the initial training phase, characteristic properties of typical image features are isolated and, based on these, a unique description of each classification category, i.e., training class, is created. In the subsequent testing phase, these feature-space partitions are used to classify image features. In supervised classification, statistical processes (i.e., based on an a priori knowledge of probability distribution functions) or distribution-free processes can be used to extract class descriptors. Unsupervised classification relies on clustering algorithms to automatically segment the training data into prototype classes. The proposed method is tested using Na Bayes Classifier, KNN Classifier and NN Classifier.

5.1 Naives Bayes Classifier

A naive Bayes classifier is a simple probabilistic classifier based on applying Bayes' theorem with strong (naive) independence assumptions. A more descriptive term for the underlying probability model would be "independent feature model". In simple terms, a naive Bayes classifier assumes that the presence (or absence) of a particular feature of a class is unrelated to the presence (or absence) of any other feature, given the class variable. Depending on the precise nature of the probability model, naive Bayes classifiers can be trained very efficiently in a supervised learning setting. Bayes classifiers have worked quite well in many complex real-world situations. An advantage of the naive Bayes classifier is that it only requires a small amount of training data to estimate the parameters (means and variances of the variables) necessary for classification.

5.2 Neural Network Classifier

Neural network based back propagation network is used for training and classification of feature matrix. Back Propagation Network is chosen since the network can adopt for noisy input and closest match can be obtained easily. Also, the error rate can be reduced considerably. Mean square error (MSE) is the error correction rule used to find the difference between the obtained output and target. The training of the network is carried on until the error reaches the predefined value given as the goal of training. Once the network is trained, it is saved as database. For creation of database, training process is done in which training parameters are fixed. The activation function for hidden and output layer is fixed as log sigmoid activation function. Error correction factor used is mean square error (‘mse’), goal for training is fixed as 1e-10 (0.0000000001) for which the network classifies the input correctly and maximum number of iterations is fixed as 10000.

5.3 KNN Classifier

The k-nearest neighbor algorithm (K-NN) is the generally used method for classifying objects based on closest training examples in the feature space. K-NN is a type of instance-based learning where the function is only approximated locally and all computation is deferred until classification. In K-NN, an object is classified by a majority vote of its neighbors, with the object being assigned to the class most common amongst its k nearest neighbors (k is a positive integer, typically small). If $k = 1$, then the object is simply assigned to the class of its nearest neighbor.

In the classification phase, the feature vector of the input fingerprint is compared with the feature vectors in the database by using the KNN classifier. The distance measure used in the classifier is ‘Euclidean Distance’. The internal database samples are divided into two sets as 2/3 for training and 1/3 for testing and gender classification is made.

6 Experimental Results

6.1 Data Acquisition

All ear images are collected internally using the Nikon Coolpix L18 colour camera. As it was decided to use the ear images to find the gender as well as the age group, samples were collected from five age groups: up to 12, 13–19, 20–25, 26–35 and 36 and above. Side face images was taken at different timings for interoperability, i.e., for better results irrespective of different light intensities. Ear images are cropped from the side ear faces and stored as internal database. To verify the robustness of the proposed algorithm, ear images are collected from the same individuals at three different timings with different lightings. Ear database is created which contains 100 sets (left and right ear) of male samples and 100 sets of female samples. For the experimentation, only good quality samples of 60 male and 60 female samples were used.

Table 1 Confusion matrix for the left ear

Actual class	Predicted class		
		Male	Female
	Male	54	6
	Female	7	53

Table 2 Confusion matrix for the right ear

Actual class	Predicted class		
		Male	Female
	Male	51	9
	Female	6	54

Table 3 Confusion matrix for the left ear

Actual class	Predicted class		
		Male	Female
	Male	50	10
	Female	3	57

Table 4 Confusion matrix for the right ear

Actual class	Predicted class		
		Male	Female
	Male	58	2
	Female	12	48

6.2 NN Classifier

The confusion matrix for the results obtained for the left ear and right ear are shown in Tables 1 and 2 respectively.

6.3 Bayes Classifier

The confusion matrix for the Bayes classifier for male and female samples is tabulated and analysed. The confusion matrix for the results obtained for the left ear and right ear are shown in Tables 3 and 4 respectively.

Table 5 Confusion matrix for the left ear

Actual class	Predicted class		
		Male	Female
	Male	58	2
	Female	8	52

Table 6 Confusion matrix for the right ear

Actual class	Predicted class		
		Male	Female
	Male	53	7
	Female	6	54

6.4 KNN Classifier

The confusion matrix for the KNN classifier for male and female samples are tabulated and analysed. The confusion matrix for the results obtained for the left ear and right ear are shown in Tables 5 and 6.

6.5 Performance Comparisons of the Classifiers

The results of all the classifiers are compared in the Fig. 6. The performance of the KNN classifier is approximately 3 % more than NN classifier and approximately 2 % more than the Bayes classifier.

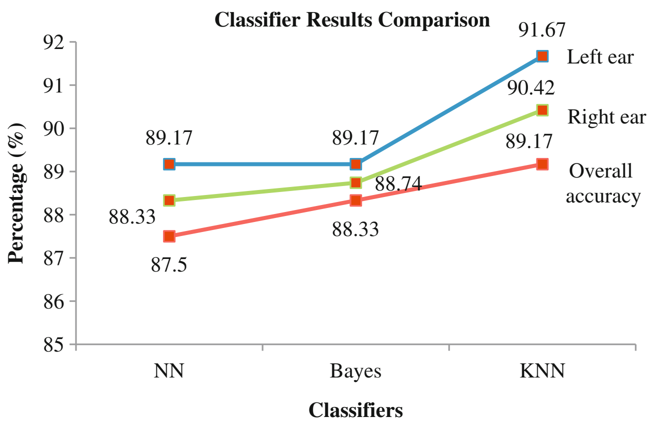


Fig. 6 Gender classification results of various classifiers

7 Conclusions and Feature Work

In this work standard features of ear biometrics has been are used for gender classification. The NN, Bayes and KNN classifiers are used to test the method. Internal database of ear pints has been used for the classification. Only 60 sets of good quality male and female samples out of 100 sets of each gender are used for classification. The quality of the ear print is affected due to the light variations, occlusion and degradation (due to motion blur). Each classifier output is discussed and compared. Of the three classifiers, KNN classifier produced good results of 91.67 % for left ear and 89.17 % for right ear. As an overall identification rate, KNN classifier accuracy rate is achieved as 90.42 %. Authors are extending their work toward providing solution for the poor quality images and partial ear prints detected at the crime scenes. Also, authors are working to find different features of ear to improve the results and to estimate the age from the ear features.

References

1. Abaza A, Ross A, Hebert C, Harrisson MAF, Nixon MS (2010) A survey on ear biometrics. In: ACM transactions on embedded computing systems, vol 9(4)
2. Rutty GN, Abbas A, Crossling D (2005) Could earprint identification be computerised? An illustrated proof of concept paper. Int J Legal Med 119:335–343
3. Jain A, Hong L, Pankati S (2000) Biometric identification. Commun ACM 43(2):91–98
4. Cummings AH, Nixon MS, Carter JN (2011) The image ray transform for structural feature detection. Pattern Recogn Lett 32(15):2053–2060
5. Alvarez L, Gonzalez E, Mazorra L (2005) Fitting ear contour using an ovoid model. In: 39th international Carnahan conference on security technology (CCST05)
6. Arbab-Zavar B, Nixon M (2007) On shape-mediated enrolment in ear biometrics. In: 3rd international symposium on visual computing (ISVC07)
7. Bustard JD, Nixon M (2008) Robust 2d ear registration and recognition based on sift point matching. In: 2nd IEEE international conference on biometrics theory, applications systems)
8. Islam S, Bennamoun M, Davies R (2008) Fast and fully automatic ear detection using cascaded adaboost. In: Proceedings of IEEE workshop on application of computer vision, pp 1–6
9. Cummings A, Nixon M, Carter J (2010) A novel ray analogy for enrollment of ear biometrics. In: Proceedings of the biometrics: theory, applications, and systems BTAS, Washington
10. Irannarelli A (1989) Ear identification. Forensic identification series. Paramount Publishing Company, Fremont
11. Lammi H (2004) Ear biometrics. Technical Report, Lappeenranta University of Technology
12. Pun K, Moon Y (2004) Recent advances in ear biometrics. In: Proceedings of the IEEE international conference on automatic face and gesture recognition, pp 164–169
13. Islam S, Bennamoun M, Owens R, Davies R (2008) Biometric approaches of 2D-3D ear and face: Advances in computer and information sciences and engineering. Springer science, pp 509–514
14. Sobhed T (2007) In advances in: computer and information sciences and engineering. Springer, Netherlands, pp 509–514
15. Victor B, Chang K, Bowyer KW, Sarker S (2002) An evaluation of the face and ear biometrics. In: IEEE international conference on pattern recognition, pp 492–43

16. Chang K, Bowyer KW, Sarkar S, Victor B (2003) Comparison and combination of ear and face images in appearance-based biometrics. IEEE Trans Pattern Anal Mach Intell 25(9):1160–1165
17. Burge M, Burger, W (1997) Ear biometrics for machine vision. In: Proceedings of the 21th workshop of the Austrian association for pattern recognition
18. Burge M, Burger W (2000) Ear biometrics in computer vision. In: Proceedings of the 15th international conference on pattern recognition ICPR, pp 826–830
19. Hurley D, Nixon M, Carter, J (005) Force field feature extraction for ear biometrics. Comput Vis Image Underst 98(3):491–512
20. Ansari S, Gupta P (2007) Localization of ear using outer helix curve of the ear. In: Proceedings of the international conference on computing: theory and applications, pp 688–692
21. Hajsaid E, Abaza A, Ammar, H (2008) Ear segmentation in color facial images using mathematical morphology. In: Proceedings of the 6th biometric consortium conference BCC, Tampa
22. Prakash S, Jayaraman U, Gupta P (2008) Ear localization from side face images using distance transform and template matching. In: Proceedings of the 1st IEEE workshops image processing theory, tools and applications IPTA
23. Prakash S, Jayaraman U, Gupta P (2009) A skin-color and template based technique for automatic ear detection. In: Proceedings of the 7th international conference on advances in pattern recognition ICAPR
24. Yan P, Bowyer K (2005) Empirical evaluation of advanced ear biometrics. In Proceedings of the IEEE computer vision and pattern recognition
25. Abaza A, Ross A (2010) Towards understanding the symmetry of human ears: a biometric perspective. In: Proceedings of the biometrics theory application system BTAS
26. Rutty G, Abbas A, Crossling D (2005) Could earprint identification be computerised? An illustrated proof of concept paper. Int J Legal Med 119:333–343
27. Lynch C (2000) Ear-prints provide evidence in court. Glasgow University News
28. Bamber D (2001) Prisoners to appeal as unique 'earprint' evidence is discredited. Telegraph Newspaper (UK). http://www.telegraph.co.uk/news/uknews/1364060/Prisoners-to-appeal-as-unique-earprint-evidenceis-discredited.html

Macro-block Mode Decision in MPEG-2 Video Compression Using Machine Learning

Vinay Kumar, K. G. Sharma and Anand Singh Jalal

Abstract Video compression currently is dominated by engineering and fine-tuned heuristic methods. Presently storing and transmitting uncompressed raw video requires large storage space and network bandwidth so compression is required. Many compression algorithms proposed to solve this type of problem. In this paper, we design a machine learning approach for the video compression using MPEG-2 codec. Various video compression techniques encode the video frames by applying inter and intra coding scheme. Video frames are divided into macro-blocks and each macro-block is encoded either by inter or by intra coding technique. It is an important issue to decide which coding technique will be applied to compress a given macro block. To solve this problem, we applied the machine learning approach in MPEG-2 video compression. We used support vector machine for the learning process and after learning any macro-block can be classified in intra or inter coding. Our experimental result suggests that use of machine learning in macro-block mode decision in MPEG-2 increases the PSNR while preserves the encoding and decoding time.

Keywords MPEG-2 · Inter and intra coding · Video compression · Machine learning · Macro-block mode decision

V. Kumar (✉) · K. G. Sharma · A. S. Jalal
Department of Computer Engineering and Applications,
GLA University, Mathura, India
e-mail: vinay.gla@gmail.com

K. G. Sharma
e-mail: hollyhoc@rediffmail.com

A. S. Jalal
e-mail: anandsinghjalal@gmail.com

Mohan S. and S. S. Kumar (eds.), *Proceedings of the Fourth International Conference on Signal and Image Processing 2012 (ICSIP 2012)*, Lecture Notes in Electrical Engineering 222,
DOI: 10.1007/978-81-322-1000-9_14,

1 Introduction

Digital video is usually available in compressed form. Neither DVDs nor digital television would be possible without video compression, because in an uncompressed format, digital video exceeds the capacities of currently affordable hardware: a typical movie in DVD quality would require 30 MB/s of constant throughput, for a total of 160 GB of storage. Compression in the multimedia domain refers to the process of reducing the quantity of data required to represent a multimedia object such as a video or an image. A Video is essentially a sequence of frames (still images) representing scenes in motion. It is minimally specified by a frame rate, frame dimensions and the total length of the video. In the uncompressed format, the video is impractical for any current date storage system or for any meaningful data transfer. The data reduction can be lossless which enables perfect reconstruction or lossy, in which there are distortions during reconstruction. Lossless compressions are more common in image domains, but for the case of video compression, lossy techniques are more common and popular due to substantial compression with notably very less distortions in the reconstruction [1].

A Video Compression encoding scheme involves two types of compression: spatial and temporal. Spatial compression focuses on compressing any given still frame or block, looking at it as an unrelated entity separated from the video. Temporal compression on the other hand uses the invariance of frames or blocks in the video over time and removes redundancy in the time domain. An actual compression scheme involves alternating layers of temporal and spatial compression [2]. There are research papers from as early as 2006 suggesting the use of Machine Learning, but none of them have adapted to commercial encoders. Current encoders still rely on heuristics and computationally expensive fine tuning of parameters, which grossly slows down the encoding process. In 1994, The MPEG-2 project was approved by extending the compression technique of MPEG-1 [3]. ISO/IEC 13818-2 standardized MPEG-2 in 2000 [4]. MPEG-2 is designed for digital television broadcasting applications that require a bit rate typically between 4 and 15 Mbps (up to 100 Mbps), such as Digital high definition TV (HDTV), Interactive Storage Media (ISM) and cable TV (CATV) [5, 6].

In H.264, the macro-block mode decision is the most computationally expensive process. The H.264 standard uses of the Rate-Distortion (RD) optimization method to choose the best macro-block mode decision which is based on a Lagrange multiplier [7] [8], whereas, in MPEG-2, macro-block mode decision is straight forward and computationally efficient but does not produces better compression. The main elements that require to be addressed in the design of an efficient MPEG-2 compression are [9]: the inter-frame prediction and the intra-frame prediction. Each one of these elements requires to be examined and various research efforts are underway. In this paper, we focus our attention on a part of the inter-frame prediction: the macro-block mode decision, one of the most stringent tasks involved in the video compression process.

The rest of the paper is structured as follows. Section 2 reviews the macro-block mode decision in MPEG-2. Section 3 introduces our macro-block mode decision algorithm for inter-frame prediction based on machine learning techniques designed for MPEG-2. In Sect. 4, we carry out a performance evaluation of the proposed algorithm in terms of its computational complexity and PSNR value. Finally, Sect. 5 draws our conclusions and outlines our future research plans.

2 Macro-block Mode Decision in MPEG-2

The MPEG-2 bit stream is basically just a series of coded frames one after the other [10]. Coding a frame in MPEG-2 format always begins by representing the original color frame in YCbCr format. The Y component represents luminance, while Cb and Cr represent chrominance differences. The three components of this color model are mostly uncorrelated so this transformation is a useful first step in reducing redundant information in the frame.

MPEG-2 uses block based coding. This means that a frame is not encoded as a whole; it is divided into many independently coded blocks. A macro-block is 16×16 pixels and is a basic unit of MPEG-2 coding. However, each macro-block is further divided into 8×8 pixel blocks. These block sizes were chosen in part because small sections of a frame of natural video (not computer generated or edited) are likely to be correlated. This correlation helps the next stages of encoding work more efficiently.

The next encoding step can vary from frame to frame. There are actually three possible types of frames, called I, P, and B frames. An I frame is intra or spatially coded so that all the information necessary to reconstruct it is encoded within that frame's segment of the MPEG-2 bit stream. It is a self contained image compressed in a manner similar to a JPEG image. A P frame is inter or temporally coded. That means it uses correlation between the current frame and a past frame to achieve compression. The MPEG-2 standard dictates that the past frame must be an I or P frame, but not a B frame. A B frame is simply a more general version of a P frame and it can refer not only to a past frame, but to a future frame, or both a past and future frame.

Encoded frame size greatly depends on the quantization scale value, but the relative frame sizes demonstrate how well temporal coding can work. In general, a P frame will be 10 % the size of an I frame, and a B frame will be only 2 % of the size of an I frame. The ordering of I, P, and B frames is fairly flexible in MPEG-2. The first frame must always be an I frame because there is no other data for a P or B frame to reference. After that, I, P, and B frames can be mixed in any order. Experimentation has shown that a repeating sequence such as I P P P B B yields good quality and compression. P frames reference the last I or P frames and the B frames reference the closest past and future I or P frames.

Figure 1 shows the procedure of macro-block mode decision in MPEG-2. All the macro-blocks of I-frame are encoded using intra coding scheme, whereas all the macro-blocks of P- and B-frames are encoded using inter coding scheme. In the next section, we treat this macro-block mode decision problem as a classification problem and use the machine learning to classify the macro-blocks of P and B frames into intra and inter macro-block.

3 Proposed Methodology

In the proposed method, we use the machine learning for the macro-block mode decision in MPEG-2. First we train the SVM (supervised learning) with some known blocks of type inter and intra and create a SVM training model. After the training we apply trained SVM to classify any input block of a P or B frame into inter or intra block. We have applied this algorithm on only P and B type of frames because I frame are encoded using intra coding method. In this section, we describe training and classification of macro-blocks of P and B frames using SVM. We define the proposed work into following step.

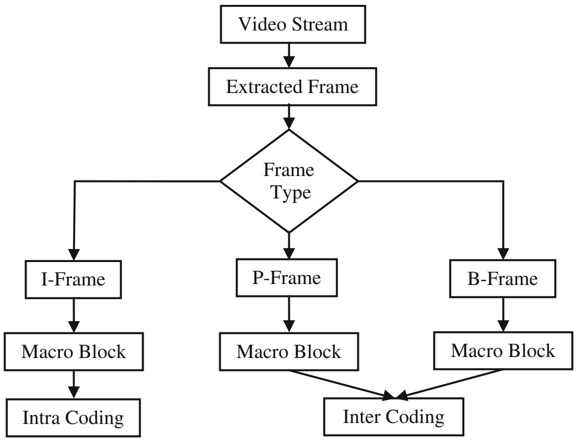
Step 1: SVM training model creation.

Step 2: Macro-block mode decision (classification) using trained SVM.

Step 3: Video compression using inter and intra coding.

The proposed framework to classify the macro-blocks is shown in Fig. 2. We encode the macro-blocks of I frames using intra coding as it is encoded in the MPEG-2, but decision for macro-blocks of P and B frames is taken using the classification by trained SVM. If the output of SVM is negative for a macro-block then that block will be encoded using intra coding otherwise inter coding.

Fig. 1 Macro-block mode decision in MPEG-2



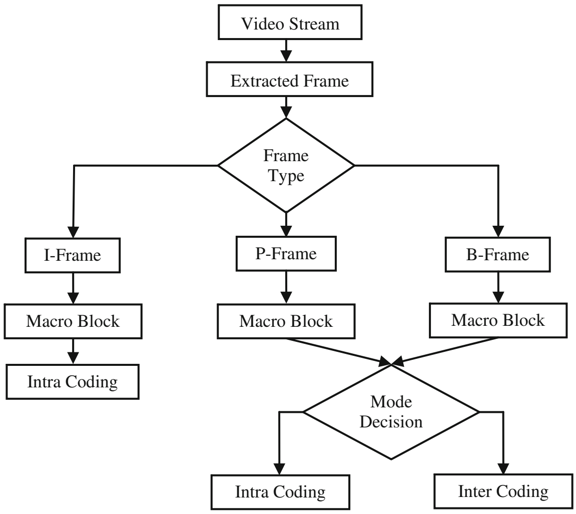


Fig. 2 Framework for macro-block mode decision using SVM as a base learner

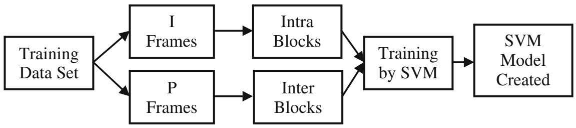
3.1 *SVM Model Creation*

For the training, we first create a sequence of macro-blocks from the I and P type of frames taken from our dataset of videos. Since we are using supervise learning so we have taken all the blocks of I frames as intra blocks and all the blocks of P frames as inter blocks. Finally we have two files: in the first file we store all the training macro-blocks while in the second file we store the corresponding coding mode. After creating those files, we train the SVM using these files and a training model will be created. Figure 3 presents the steps of SVM model creation using macro-blocks of type I and P frames. After extensive experimentation, we found that sequences that contain regions varying from homogenous to high-detail serve as good training sets. Good sample sequences could be Flower and Football.

3.2 *Classification*

In this step, we perform a macro-block mode decision algorithm aiming to increase the quality of compressed video. We can classify the macro-blocks of P and B type of frames of any input video into intra and inter blocks. In the Fig. 4, it is clear that after trained the SVM we classify the macro-block into inter or intra using SVM model.

Fig. 3 Process of training using SVM base learner



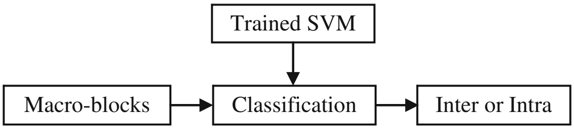


Fig. 4 Macro-block mode decisions (classification) using trained SVM

3.3 Intra and Inter Coding

Both intra and inter coding involves the concept of Discrete Cosine Transform (DCT). Each block of the frame is processed independently with DCT. This transform generates a representation of each block in the frequency domain instead of the spatial domain. Intra coding uses only the normalized DCT of a block of current frame only, whereas inter coding uses the normalized DCT of a block as well as the normalized DCT of corresponding block from past and future frames.

4 Results and Discussion

In this section, first we discuss about the video data set which is used in this paper to verify our proposed method. Then we show a detailed result analysis and discuss several issues regarding the performance of the proposed approach and compare our result with the MPEG-2.

4.1 Data Set

To demonstrate the performance of the proposed approach, we have used a standard video data set (taken from http://media.xiph.org/video/derf/). We have used following 15 videos from the derf's collection to evaluate our program: akiyo.y4m, bowing.y4m, bus.y4m, coastguard.y4m, container.y4m, flower.y4m, football.y4m, foreman.y4m, galleon.y4m, husky.y4m, mobile.y4m, news.y4m, silent.y4m, tempete.y4m, and waterfall.y4m. The resolution of all videos is 352×288 pixels. Since all videos are in the YUV format so first we convert them into RGB format and after compression will take place on the RGB format..

4.2 Experimental Results

In this sub-section, we show the detailed result analysis and discuss the performance of the proposed approach in comparison with MPEG-2. As examples of the simulation's capabilities, we encoded and decoded 150 frames of each movie of the data set with the quality scale ranging from 1 to 128 and calculated the encoding time, decoding time and PSNR against the scaling parameter. The training time is excluded from the encoding and decoding time because we are doing the training before the compression.

For starting 4 frames of movie football.y4m, the original, reconstructed, and error images are shown in the Fig. 5 with the quality scale set to 64. Note that the error shows positive and negative error as differences from a middle reference frame (indicates no error). The middle reference frame is shown as the first figure in Fig. 5. From this Figure it is clear that the reconstructed image have got less distortion using machine learning approach while the value of scaling parameter is so high. In Fig. 6, we show the effect of scaling parameter on the first frame of the movie akiyo.y4m considering different value of scaling parameter using the machine learning in MPEG-2.

Thus, on a lower scale there will not be any difference between the quality of the original frame and the frame reconstructed. But as this scale will increase, the

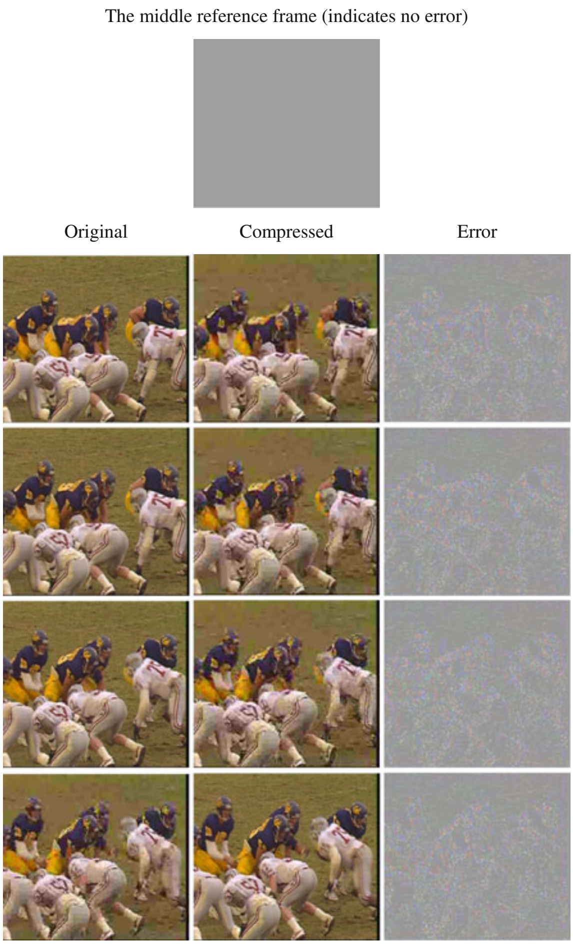


Fig. 5 Original, reconstructed, and error images of starting 4 frames of movie football.y4m when scaling parameter is set to 64

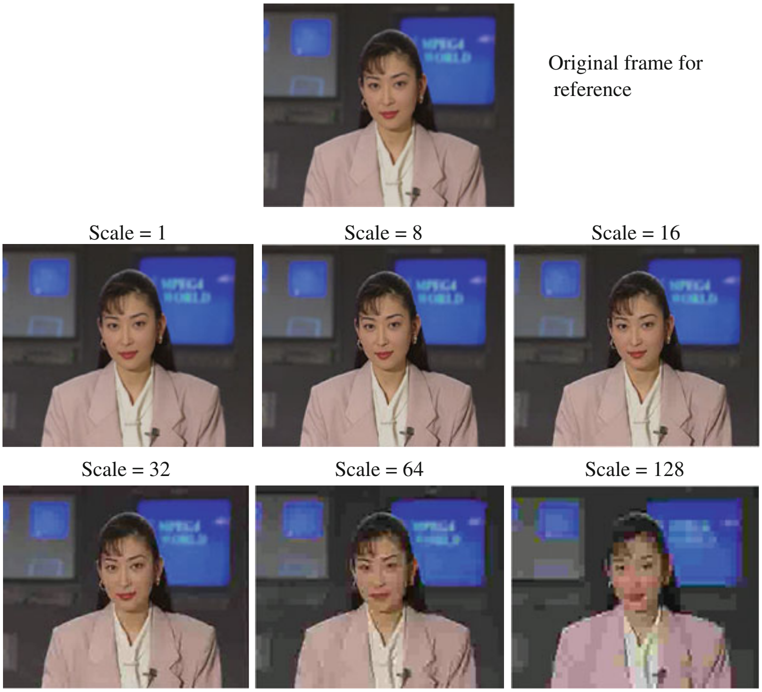
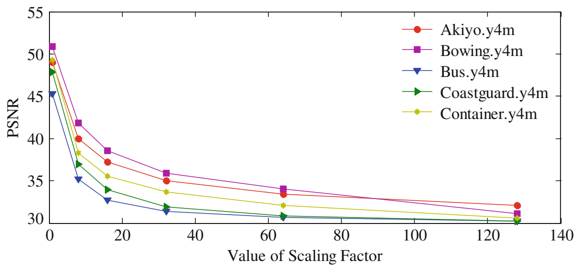


Fig. 6 Original and reconstructed images of first frame of movie akiyo.y4m for the different scaling parameters using machine learning approach

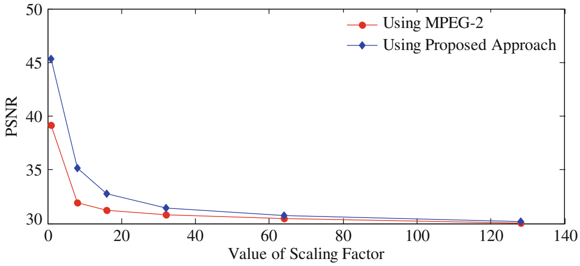
Fig. 7 PSNR between five original and its corresponding compressed video sequences considering different values of quality scaling using proposed method



quality of the frame will decrease. On the scale 8, there will be a little difference visible on the sides of the frame while on scale 128, the frame will become blurred. All these changes happen because of the color encoding in the frame.

Figure 7 shows the PSNR between the original and compressed video sequences of Akiyo, Bowing, Bus, Coastguard and Container on the basis of different quality scaling using SVM for the macro-block mode decision in MPEG-2. For all the video sequences the PSNR increases with decreasing the value of the scaling factor. Among these video sequences, Bowing shows better quality because the presence of more smooth areas and less edging regions in its frames.

Fig. 8 Comparison of PSNR achieved using MPEG-2 and proposed approach on the bus video sequence considering different values of quality scaling



We have compared our experimental result with MPEG-2 in the Fig. 8. In this Figure, the value of PSNR is plotted against the quality scaling using MPEG-2 and proposed approach for the Bus video sequence. It shows that the PSNR achieved using proposed methodology is more than the PSNR achieved using MPEG-2 for all scaling values because our method encodes the macro-block of P and B type of frames either by intra or inter coding depending upon the output of SVM, whereas MPEG-2 encodes the macro-block of P and B type of frames only by inter coding.

Table 1 illustrates the performance of introduced method and MPEG-2 in terms of Encoding time(in second), Decoding time(in second), and PSNR on the some videos of derf's collection. We have observed that the quality of compressed video is better if we use SVM for the macro-block mode decision of P and B type of

Table 1 Encode time, decode time and PSNR values for the MPEG-2 without machine learning and MPEG-2 using machine learning (SVM) for all 15 video sequences of the data set while quality scaling is set as 8

S. No.	Video sequence name	MPEG-2			Using machine learning		
		Encode time	Decode time	PSNR	Encode time	Decode time	PSNR
1.	Akiyo	56	23	38.71	54	23	39.98
2.	Bowing	57	23	37.76	54	22	40.79
3.	Bus	56	23	31.33	54	23	35.59
4.	Coastguard	57	23	32.21	55	23	36.70
5.	Container	56	23	35.89	54	23	38.52
6.	Flower	57	24	31.85	54	23	35.88
7.	Football	56	23	32.39	53	23	36.83
8.	Foreman	56	23	34.79	54	23	38.22
9.	Galleon	56	23	33.95	56	23	36.40
10.	Husky	56	23	29.43	54	23	33.54
11.	Mobile	56	23	30.32	54	23	33.62
12.	News	56	23	36.35	54	23	37.89
13.	Silent	56	23	35.04	53	23	37.13
14.	Tempete	56	23	31.73	53	23	34.60
15.	Waterfall	56	23	32.89	53	22	34.45

frames in MPEG-2. The encoding and decoding time is nearly equal in both cases because we classify all the macro-blocks at the beginning using trained SVM, which is computationally efficient.

5 Conclusion

MPEG-2 uses intra coding for the macro-blocks of I frame and inter coding for the macro-blocks of P and B frames. In this paper, we have used both intra and inter coding for the macro-blocks of P and B frames and intra coding for the macro-blocks of I frame. We have used the concept of machine learning to decide intra or inter coding for a given macro-block of P and B frames. To implement machine learning in the scheme of thing, first we have created a SVM model for the training and then used this trained SVM for the classification of the macro-blocks of P and B frames. We have tested our approach on the derf's collection of video sequences. Our experimental results suggest that the proposed approach is able to produce a better quality of compressed video while preserving the encoding and decoding time.

References

1. Lampert CH (2006) Machine learning for video compression: macro-block mode decision. In: the Proceedings of the 18th IEEE international conference on pattern recognition, pp 936–940
2. Richardson IE (2003) H.264 and MPEG-4 Video Compression: Video Coding for Next-Generation Multumedia. John Wiley & Sons, New York
3. Abomhara M, Khalifa OO, Zakaria O, Zaidan AA, Zaidan BB, Rame A (2010) Video compression techniques: an overview. J Appl Sci 10(16):1834–1840
4. International Standards Organization (2000) MPEG2. Information technology—generic coding of moving pictures and associated audio information. ISO/IEC 13818-2
5. Sikora T (1997) MPEG digital video-coding standards. MPEG digital video-coding standards. IEEE Trans Signal Process 14:82–100
6. Puri A, Chen X, Luthra A (2004) Video coding using the H.264/MPEG-4 AVC compression standard. Signal Process Image Commun 19:793–849
7. Sullivan G, Wiegand T (1998) Rate-distortion optimization for video compression. IEEE Signal Process Mag 15(6):74–90
8. Wiegand T, Schwarz H, Joch A, Kossentini F, Sullivan GJ (2003) Rate-constrained coder control and comparison of video coding standards. IEEE Trans Circuits Syst Video Technol 13(7):688–703
9. Kalva H (2004) Issues in H.264/MPEG-2 video transcoding. In: Proceedings of the consumer communications and networking conference
10. Li ZN, Drew MS (2004) Fundamentals of multimedia, ISBN: 0130618721, Pearson Education, Inc., Upper Saddle River, NJ, 07458

Stochastic Resonance and Mean Shift Filtering for Detecting Weak Features in Noisy Images

J. V. R. Sagar and Chakravarthy Bhagvati

Abstract Stochastic Resonance has been shown to occur in many biological, physical and geological systems, resulting in the boosting of weak signals to make them detectable. In the image processing domain, narrow regions, small features and low-contrast or subtle edges, especially in noisy images, correspond to such weak signals. We show, both mathematically and empirically, that stochastic resonance occurs and may be exploited in the detection, extraction and analysis of such features. These mathematical results are confirmed by simulation studies. Finally, results on standard images such as cameraman, boats, lena, etc. demonstrate that several subtle features lost by the application of robust techniques such as Mean Shift filter are recovered by stochastic resonance. These results reconfirm the mathematical and simulation findings.

Keywords Stochastic resonance · Robust techniques · Mean-shift filter

1 Introduction

Noise has been extensively studied by both signal and image processing communities. One of the most successful approaches is the application of robust statistics starting in the 1990s. Shawe-Taylor et al. [1] state that robust techniques are those

J. V. R. Sagar (✉) · C. Bhagvati
Dept. of Computer and Information Sciences, University of Hyderabad,
Hyderabad 500046, India
e-mail: sagarjvr@yahoo.com

C. Bhagvati
e-mail: chakcs@uohyd.ernet.in

Mohan S. and S. S. Kumar (eds.), *Proceedings of the Fourth International Conference on Signal and Image Processing 2012 (ICSIP 2012)*, Lecture Notes in Electrical Engineering 222,
DOI: 10.1007/978-81-322-1000-9_15,

able to handle noisy data and approximate patterns. In particular, such techniques should be able to tolerate noise in that the output from a noisy image (or signal) should not be too different from that of a clean image. M-estimators [2, 3], RANSAC [4] and MINPRAN [5] are some early examples. Mean-Shift filter [6, 7], another robust algorithm, proved extremely popular in the early part of the last decade and many algorithms today use it as a first stage before improving its output with other methods [15].

If we consider a noisy image as one with a low signal to noise ratio (SNR), then there are two broad categories of noisy signals. The first is that the signal or the feature of interest, is strong but there is a significant amount of noise. In other words, if we were to obtain somehow a clean image, it should be possible to extract the features of interest using even a non-robust technique. The second case of low SNR is when the signal is weak. This occurs when the feature of interest has a low contrast with its neighbours making it difficult if not impossible to detect. It may also occur when the size of the feature is small compared to the kernel and therefore is not detected. Such a case is responsible for the failure of a smoothing or a median filter to preserve edges (narrow features). Note that these failures occur even if there is minimal amount or no noise.

We show in this paper that the mean-shift filter, an example of a robust technique, performs poorly in detecting weak features. Noise, especially, leads to failure. This is a major problem because mean-shift filter is designed to work on noisy images.

Our solution to the problem is *stochastic resonance*, a phenomenon responsible for boosting weak signals observed primarily in biological and physical systems [8]. We show in this paper through simulation studies that stochastic resonance occurs and boosts narrow regions and low contrast signals above the detectability threshold. Stochastic resonance and mean-shift filter together form a powerful combination for handling high-noise as well as weak-signal features. Finally, we show results on standard images such as cameraman, boats, lena, peppers and others that stochastic resonance improves the performance of mean-shift filter in detecting edges and regions.

The rest of the paper is organized as follows. Section 2 introduces stochastic resonance and certain theoretical results. Section 3 describes our empirical analysis and simulations based on the analysis presented in Sect. 2. Section 4 presents the results on standard images combining mean-shift filter and stochastic resonance while Sect. 5 presents a brief analysis and conclusions drawn from the results.

2 Stochastic Resonance

Stochastic Resonance (SR) was a term coined in 1981 by Roberto Benzi et al. in their classic papers [10–12] on recurrent ice ages. It refers to a seemingly counter-intuitive observation in several biological, geological and physical systems of the use of noise to cancel certain effects of noise and boost weak signals. In particular, it was found that adding a moderate amount of noise enables weak signals, hitherto lying below the detection threshold, to be detected. The first direct evidence of SR was reported by Fauve and Heslot [13] in 1983 through their experiments on an ac-driven Schmitt trigger. Later, McNamara et al's paper [14] on bistable ring lasers led to a series of papers by the Physics community. Today, it may not be an exaggeration to say that the Physics community is taking the lead in studies on SR. An excellent review of the early work and detailed experiments may be found in the very long 1998 survey paper [8] by Gammaitoni et al. Many nonlinear systems, such as earth climate dynamics, electronic circuits, information capacities of non-linear channels, exhibit SR effect too. Others such as [16, 17] and [18] also provide significant information on SR. As far as image processing community is concerned, Jha et al. present a cursory treatment on the application of SR to de-noising [9].

Consider a signal that lies entirely below a detection threshold and therefore is completely undetected. Adding moderate amounts of noise to such a signal causes the combination of signal and noise to cross the detection threshold without too much distortion of the characteristics of the original signal. With greater amounts of noise, the output becomes entirely dominated by noise characteristics and the signal is again undetectable. Therefore, SNR is low for both no noise and addition of large amounts of noise. SNR initially decreases until noise begins to dominate the signal. Thus, a plot of SNR versus added noise intensities shows an upside-down 'U' shape. The presence of such a curve is often taken as the evidence for SR in a system.

Several interesting systems that exhibit SR have been studied by Greenwood et al. in [19]. We follow the analysis presented there to show the existence of SR in image domain where weak signals are equated with narrow regions and low-contrast edges.

Consider a signal $s(t)$ and a detectability threshold a. The signal is detected if at any t_i, $s(t_i) \geq a$. Using a kernel k for estimating s, the minimum value of *asymptotic mean average square error* (AMASE) is shown in [19] to be

$$AMASE = \frac{5}{4} n^{-\frac{4}{5}} \left(\mu_2(K)^2 \frac{1}{n} \sum_{i=1}^{n} \frac{p(t_i)^2}{f(F^{-1}(p(t_i)))^2} \right)^{\frac{1}{5}} \left(R(K) \frac{1}{n} \sum_{i=1}^{n} \frac{p(t_i)(1-p(t_i))}{f(F^{-1}(p(t_i)))^2} \right)^{\frac{4}{5}} \quad (1)$$

where n is the number of samples, $\mu_2(k), R(K)$ are kernel constants and $p(t)$ is the kernel estimator. AMASE is the theoretical measure of the mean squared error commonly used as a goodness-of-fit estimate in model fitting. The general approach is to take a Taylor approximation of the least squared error [20]. A number of kernels—triangular, bi-weight, tri-weight, Epanechnikov—are popular in the vision community and we use the Epanechnikov kernel in this paper. The Epanechnikov kernel (more details may be found in [1]) is given by $K(u) = 3/4(1-u^2), |u| \leq 1$ and is radially symmetric. If we assume Normal error distribution and the Epanechnikov kernel, AMASE has been shown as

$$AMASE = \frac{5}{4} n^{-\frac{4}{5}} \left(\frac{1}{25n} \sum_{i=1}^{n} \left(\frac{a - s(t_i)}{\sigma^2} s'(t_i)^2 + s''(t_i) \right)^2 \right)^{\frac{1}{5}} \left(\frac{3}{5n} \sum_{i=1}^{n} \frac{\sigma^2 \phi\left(\frac{s(t_i)-a}{\sigma}\right) \phi\left(\frac{a-s(t_i)}{\sigma}\right)}{\psi\left(\frac{s(t_i)-a}{\sigma}\right)^2} \right)^{\frac{4}{5}} \tag{2}$$

In Eq. 2, $\mu_2(k) = 1/5$ and $R(K) = 3/5$ and σ^2 is the noise variance that results in SR. s' and s'' are the first and second derivatives of s. We model the weak edges and narrow regions as step edges smoothed by a Gaussian and a narrow pulse respectively.

3 Simulation Studies

In this section, we show through simulation that weak edges when modelled by a smoothed step edge exhibit SR. A weak edge is modelled by an ideal step edge that is smoothed by a Gaussian filter and we write it as

$$s(x) = A \frac{1}{\sqrt{2\pi}\sigma} \int e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \tag{3}$$

where A is the edge magnitude and μ is the location of the edge. σ is the parameter used to represent the distortion in the ideal step edge caused by the imaging system. This approach is along the lines of that followed by John Canny [21]. The first and second derivatives are given by Eqs. 4 and 5.

$$s'(x) = A \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{4}$$

$$s''(x) = A \frac{\mu - x}{\sqrt{2\pi}\sigma^3} s'(x) \tag{5}$$

It is important to note that there are two σs now: the first is the σ present in Eq. 2 and characterizes the added noise for SR; and, the second is the σ of the signal and its derivatives (Eqs. 3–5). This second σ that describes the signal is not present explicitly in Eq. 2 and is really because of the weak-edge model of Eq. 3.

Consider a signal given by Eq. 3 with $A = 1.0, \mu = 0$ and $\sigma = 1.0$. The detectability threshold $a = 1.05$ which is slightly above the maximum value of the signal making the signal undetactable. A plot of AMASE (Eq. 2) for various values of added noise with σ ranging from 0.1 to 1.6 is shown in Fig. 1a. The AMASE values are normalized to the range 0–1. The plot shows the characteristic minimum in AMASE value at $\sigma = 0.6$ indicating the occurrence of SR.

The theoretical analysis is confirmed by an empirical study. The weak step edge signal is given by the same equation (Eq. 3). The threshold is set to the same value of $t = 1.05$. Note that the signal is *perfectly* estimated if the strength of the combined signal does not exceed the threshold for all samples that are to the left of the step edge; and, exceeds the threshold for all samples to the right of the step edge. We, therefore, measure the error as the number of samples that are detected above the threshold and are to the left of the edge location; and the number of samples that lie below the threshold to the right of the edge location. Initially, the

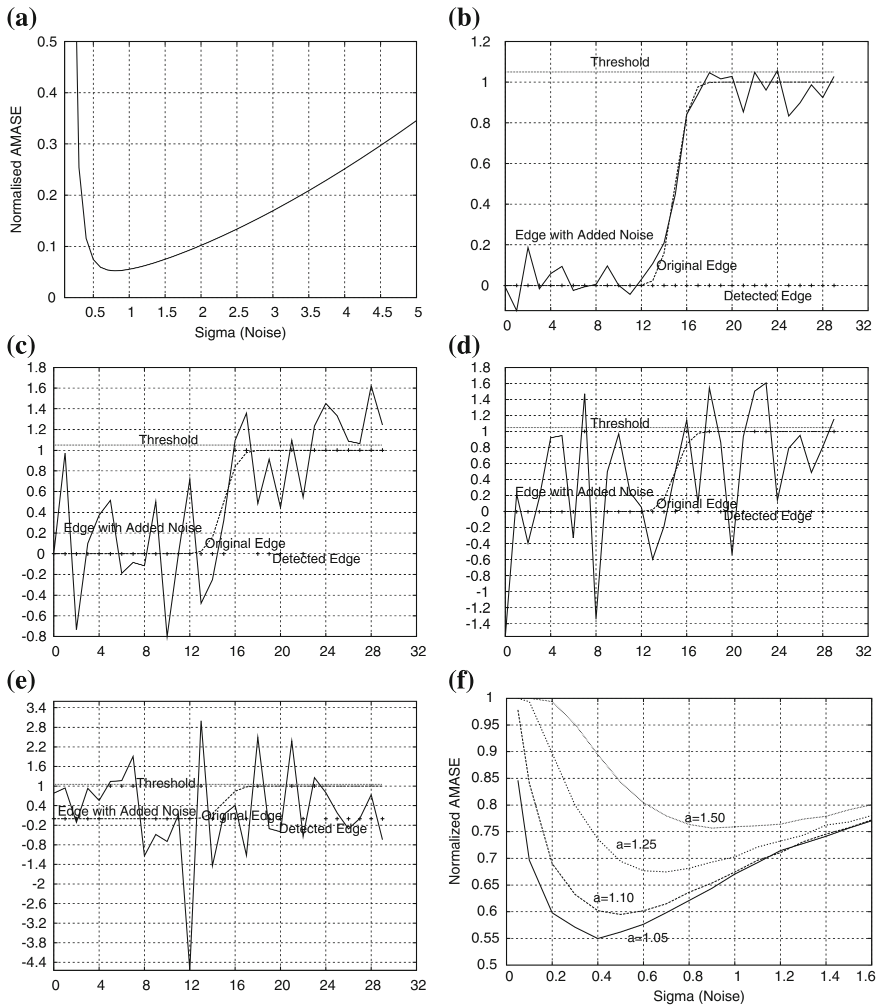


Fig. 1 Illustration of stochastic resonance on the weak edge model. **a** Theoretically calculated AMASE as a function of noise sigma (Eq. 3), **b**, **c**, **d**, **e** Weak edge corrupted with different amounts of noise ($\sigma = 0.1, 0.4, 0.75, 1.5$). **f** AMASE values calculated from the noise corrupted signals in **b**–**e**. X-axis is the sample number and Y-axis is the signal amplitude for **b**–**e**

error is 50 % because the step edge does not exceed the threshold at all and there are as many samples to the left of the edge position (and therefore are correctly not detected) as there are to the right of the edge location (which are incorrectly not detected). Then, Gaussian noise with σ ranging from 0.1 to 1.6 is added to the signal.

Figure 1b–e illustrate the empirical study. The step edge corresponding to Eq. 3 is labelled *Original Edge* in the plots. *Edge with Added Noise* indicates the combined signal while the threshold value is shown as the horizontal line labelled *threshold*. If the value of the signal exceeds the threshold, then it is detected as a 1, otherwise it is undetected and has a value 0. *Detected Edge*, indicated by the '+' symbols, is the result of thresholding. It may be seen in Fig. 1b, with the addition of noise having $\sigma = 0.1$, only one sample (No. 24) is detected, i.e., exceeds or equals the threshold. Thus, the error is 46.67 % (14/30), that is 14 out of the 30 samples are incorrectly estimated. Note that this is a small improvement over the initial error of 50 %. As we increase the added noise to $\sigma = 0.4$, many more samples to the right of the step edge cross the threshold and are detected (Fig. 1c). As we further increase the noise to $\sigma = 1.5$, many samples both to the left and right of the edge location cross the threshold and the error, consequently, increases. If the theory is correct and SR occurs, then there is a specific level of noise for which the error is minimised. A plot (Fig. 1f for $t = 1.05$) of the error versus σ (noise) confirms SR and shows the same characteristic as that of the theoretical AMASE curve of Fig. 1a. The minimum, however, appears at $\sigma = 0.45$ which is slightly lower than the theoretical value of $\sigma = 0.6$ which may be due to the discrete nature of the sampling in the empirical study. Note that the number of samples is set to 30 for the plots so that the analysis is clear. In the calculations shown in Fig. 1f, the number of samples is 10000 and the experiment is repeated 20 times for each case to get more accurate estimates of the average errors.

Other plots in Fig. 1f are for different values of the detectability threshold a. As the value of a increases it indirectly makes the edges weaker and harder to detect. However, SR occurs in all the cases but we need to add *greater* amounts of noise, indicated by the increased values of σ (SR noise) at which the minima occur, to boost the increasingly weak edge strengths. The minima are also at higher values of AMASE which suggests that the improvement reduces with increasingly weak edge strengths (higher values of a). Also, the valleys are becoming shallower and it suggests that when the edges are extremely weak, SR may not help. In fact, when the threshold $a = 10$, it has been observed that there is no SR in that the AMASE curve does not show any minimum.

Similar theoretical and empirical analysis has been done for narrow regions which are modelled as a pulse of width $2a$ and height h. The smoothing is modelled as two different step edges—one going from low to high, and the other, from high to low—separated by $2a$. The results for such a case also show that SR occurs. We are not presenting the details in this paper in the interest of brevity.

4 Experiments on Noisy Images

In this section, we show that mean-shift filter can be combined with SR to reveal weak edges and other features that are missed by applying mean-shift filter alone. Mean-shift filter [7] is an example of a non-parametric kernel density based technique for robust segmentation. Mean Shift analysis applies kernel density estimation to segmentation and edge detection of images. An image is represented as a 2D lattice of one-dimensional vectors resulting in a 3D feature space for grayscale and a 5D space for color images. The local structure is explored by dividing the feature space into cuboids called buckets. The size of the bucket is controlled by two parameters σ_s and σ_r along the spatial and the intensity (range) axes. For any pixel x, the size of the local neighborhood and distribution are given by the choice of a kernel function $k(x)$. We use the Epanechnikov kernel and 27 buckets $(3 \times 3 \times 3)$ surrounding the bucket containing the pixel under consideration for estimating the kernel density gradients. The successive application of mean shift defines a path which leads to the local density maximum. In the mean shift filtered output, all the points that lie on the path are assigned the local density maxima, i.e. *modes*, as their values. This description is substantially same as the one given in [7].

The outputs from mean-shift algorithm for three images corrupted by the addition of Gaussian noise of 10, 20 and 30 % (Fig. 2a–c) are shown in Fig. 2d–f. For relatively low noise levels of 10 and 20 %, mean-shift filter preserves weak edges and narrow regions. In Fig. 2d, the two sets of three narrow masts each on top of the boat are clearly seen. There are also low-contrast edges on the tower to which the masts are fixed. These are also clearly seen. In Fig. 2e, with noise of 20 %, the masts are 'broken' and unclear, while in Fig. 2f, the masts are no longer visible and even the tower to which they are attached is so highly distorted in shape that it is virtually unrecognizable. These results support our hypothesis that robust techniques render undetectable certain features that are detectable in the original images.

We applied SR to the mean-shift filtered output shown in Fig. 2d–f. As we know the parameters of the buckets used in the mean-shift algorithm, we know the size of the neighbourhoods that participated in deriving the mean-shift. We used $\sigma_s = 4$ and $\sigma_r = 8$ for the highly noisy images and from the size of the buckets, we estimated the optimal noise sigma as $\sigma = 1.2$. It is done by using experiments similar to the ones described in Sect. 3 with varying values of σ for smoothing the step edges. These σ are analogous to the size of the buckets. We then added the optimal amount of noise to the mean-shift filtered output and thresholded each pixel using a local threshold. The local threshold is the average gray scale value in the neighbourhood of the pixel. The final images so generated for the images in Fig. 2d–f are shown in Fig. 3a–c.

It may be seen from Fig. 3a, that at low-noise levels, SR results in a cleaner image but appears to remove some low-contrast edges. Note the differences in the appearance of the tower on the boat. However, the narrow regions are preserved, if

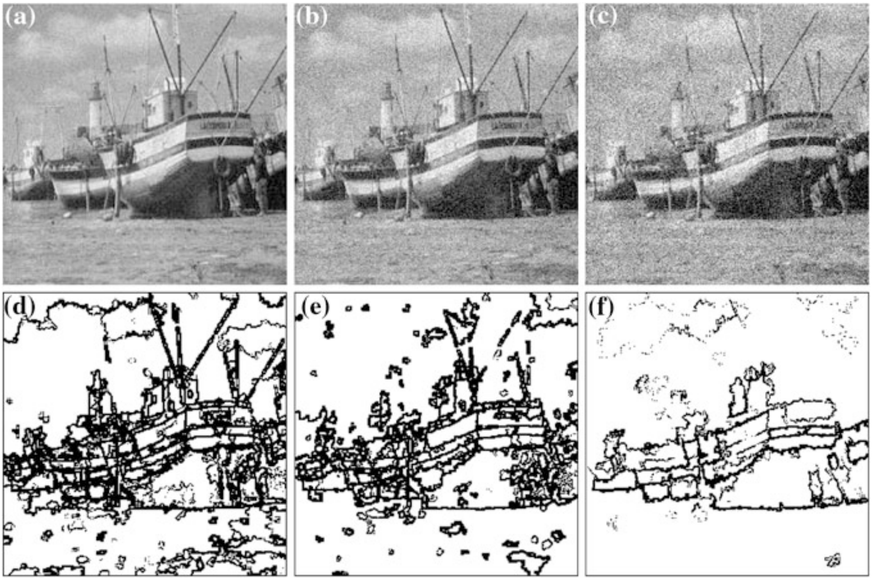


Fig. 2 Performance of mean-shift filter. Input images with **a** 10 %, **b** 20 % and **c** 30 % noise. Corresponding outputs are in **d**, **e** and **f** respectively

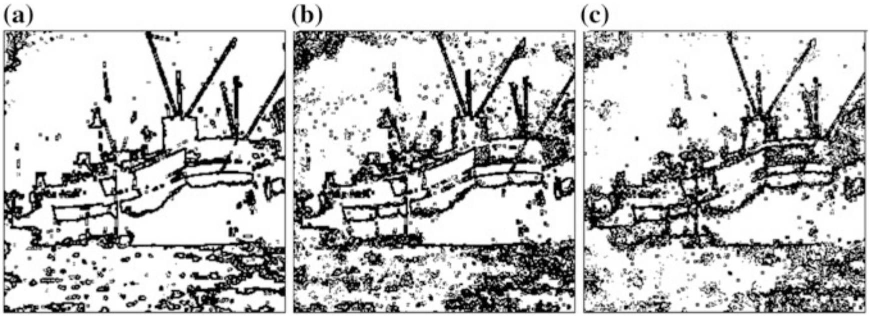


Fig. 3 Filtered outputs for images shown in Fig. 2 **a–c** after combining mean-shift filter with Stochastic Resonance

not improved. For higher noise levels (Fig. 3b, c), it may be seen that the outputs continue to preserve and in the case of of Fig. 3c, even re-detect the narrow masts which were undetected in the original mean-shift output. The tower, which has become distorted in Fig. 2f, is also better detected. Thus, the addition of SR to mean-shift filter combines the tolerance to noise for gross features that mean-shift filter has with the ability to detect weak and narrow regions that SR gives. The result is noisy and there is a need to perform post-processing to clean up the image.

We have done experiments on doing SR directly on the 30 % noise corrupted input image, but the image, as expected, is noisier than the one obtained when using mean-shift.

We tested the combination of SR and mean-shift filtering on a set of 40 images corrupted by 10, 20 and 30 % Gaussian noise (giving a set of 120 images) and found that the combination improves the performance of either if applied individually. The test image set includes several standard images such as cameraman, peppers, lena, etc. In all these cases, the improvement is in the detection of the weak and narrow regions which were missed by the mean-shift algorithm. The result is more noisy than the mean-shift filtered output, but for applications where the weak and narrow region detection is important, SR provides a powerful tool.

5 Discussion and Conclusions

The results on images with varying levels of noise lead us to believe that Stochastic Resonance may be brought into the Computer Vision and Image Processing domain as a noise handling paradigm. Robust techniques have demonstrated their utility in handling noise but with the requirement that the signal be detectable. The results presented in this paper and by others earlier show that noise levels up to 20 % can be handled by robust techniques. When the noise levels are higher or if the signal is weak, then the robust techniques fail to produce the desired results. Our results show that Stochastic Resonance works even on images with 30 % and higher noise as well as on weak, barely detectable signals. Thus, it is possible to state that Stochastic Resonance may be complementary to robust techniques in that robust techniques are useful in dealing with low SNR signals caused by high noise (with S strong enough to be detectable above a threshold) while SR improves low SNR caused by weak signals. In Physics and other fields, several problems involving complex fields and potential well structures have been successfully solved using SR with profound applications in quantum mechanics, solid state physics, etc.; whether such procedures and solutions have a similar impact on computer vision and image processing remains to be seen.

In this paper, we proposed a novel paradigm for handling noisy images in computer vision applications. The novelty lies in combining certain robust techniques, in particular mean shift filtering, with stochastic resonance. The results obtained on 120 images show that the SR and mean-shift work in a somewhat complementary fashion to preserve weak features while handling noisy images. Stochastic resonance improved upon mean shift filter in the case of low SNR caused by high noise when the underlying signal is strong enough to be detectable. The authors are convinced that the study of SR in computer vision merits greater attention.

References

1. Taylor JS, Cristianini N (2004) Kernel methods for pattern analysis. Cambridge University Press, New York
2. Hampel FR, Rousseeuw PJ, Ronchetti E, Stahel WA (1986) Robust statistics: the approach based on influence functions. Wiley, New York
3. Huber PJ (1981) Robust statistics. Wiley, New York
4. Fischler MA, Bolles RC (1981) Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Comm ACM 24(6):381–395
5. Stewart CV (1995) MINPRAN: a new robust estimator for computer vision. IEEE Trans Pattern Anal Mach Intell 17:925–938
6. Comaniciu D, Meer P (1999) Mean shift analysis and applications. In: Proceedings of the 7th IEEE international conference on computer vision (ICCV99), vol 2, pp 1197–1203
7. Comaniciu D, Meer P (2002) Mean shift: a robust approach toward feature space analysis. IEEE Trans Pattern Anal Mach Intell 24(5):603–619
8. Gammaitoni L et al (1998) Stochastic resonance. Rev Mod Phys 70(1):223
9. Jha RK, Biswas PK, Chatterji BN (2005) Image denoising using stochastic resonance. In: Proceedings of the international conference on cognition and recognition, Mysore
10. Benzi R, Sutera A, Vulpiani A (1981) The mechanism of stochastic resonance. J Physics A 14:L453
11. Benzi R, Parisi G, Sutera A, Vulpiani A (1982) Stochastic resonance in climatic change. Tellus 34(1):10–16
12. Benzi R, Sutera A, Parisi G, Vulpiani A (1983) A theory of stochastic resonance in climate change. SIAM (Soc Ind Appl Math) J Appl Math 43:565
13. Fauve S, Heslot F (1983) Stochastic resonance in a bistable system. Phys Lett 97A:5
14. McNamara B, Wiesenfeld K, Roy R (1988) Phys Rev Lett 60:2626
15. Winbing Tao YZ, Jin H (2007) Color image segmentation based on mean shift and normalized cuts. IEEE Trans Syst Man Cybern 37:1382–1389
16. Wiesenfeld K, Wellens T, Buchleitner A (2002) Coherent evolution in noisy environments. Springer, Berlin
17. Wellens T, Shatokhin V, Buchleitner A (2004) Stochastic resonance. Rep Prog Phys 67(1):45–105
18. Pascual JC, Ordonez JG, Morillo M (2005) Stochastic resonance: theory and numerics. Chaos 15:1–12
19. Greenwood PE, Muller UU, Ward LM, Wefelmeyer W (2003) Statistical analysis of stochastic resonance in a thresholded detector. Austrian J Stat 32(1,2):49–70
20. Müller UU (2000) Nonparametric regression for threshold data. Canadian J Stat 28:301310
21. Canny JF (1986) A theory of edge detection. IEEE Trans Pattern Anal Mach Intell 8:147–163

Integration of Camera Systems for Determining Altitude of Low-Flying Aircraft Above Water

Rao Vandana Parankusam and Nebylov Alexander V.

Abstract Low altitude aircraft that fly above water especially (the Russian) Ekranoplanes are finding wide applications not only as a fast and viable transportation means over various water bodies, but also as rescue vehicles for critical rescue operations. Their low-altitude flight makes it necessary for them to be equipped with an accurate altimeter which distinguishes even minute changes easily. Computer vision systems are finding applications for ranging, mapping and object recognition in various fields. Two separate systems that used the principles of 'computer vision'—a passive stereoscopic camera system (PSCS) and an active stereoscopic camera system (ASCS) were proposed for measuring altitude in earlier papers. Both these systems are explained briefly in this paper. PSCS operates only in rough sea conditions, whereas ASCS works only in calm sea conditions. Thus, individually they cannot operate in all-sea states. This paper proposes a way to integrate the camera systems so as to determine altitude in all sea states.

Keywords Active camera system · Passive camera system · Altitude measurement · Computer vision

R. V. Parankusam (✉) · N. A. V.
St. Petersburg State University of Aerospace Instrumentation, St. Petersburg, Russia
e-mail: pvandanaraox@yahoo.com

N. A. V.
e-mail: nebylov@aanet.ru

Mohan S. and S. S. Kumar (eds.), *Proceedings of the Fourth International Conference on Signal and Image Processing 2012 (ICSIP 2012)*, Lecture Notes in Electrical Engineering 222,
DOI: 10.1007/978-81-322-1000-9_16,

1 Introduction

Ekranoplanes work on the principle of Ekrano effect. Ekrano effect is the dynamic cushion of air formed under the wings of an aircraft, generally noticeable during take-off and landing. This effect is also called 'wing-in-ground' effect. Such aircraft have very low altitude of flight—just 5 m above water as shown in Fig. 1.

Different kinds of Ekranoplanes were developed which weighed from several hundred kilograms up to 500 tons. Heavy Ekranoplanes may be equipped with radio altimeters and other expensive and complex sensors of motion parameters, but for small Ekranoplanes very simple, accurate, and cheap sensors are necessary. Presently, most Ekranoplanes do not have any altitude sensors and automatic control systems at all. It is necessary to develop new concepts of designing such sensors.

The Ekranoplane under consideration here is the "spasatel" or "the rescuer" whose cruise altitude is just 5 m above sea level, has a cruising speed of 400 km/hr and operates in sea-states from 0 to 5.

It is important to note that Ekranoplanes do not cause any additional disturbances of the sea surface, after their take-off stage is over and before their landing stage. [1, 2].

Computer vision systems hold the potential of not only working as altimeters, but also as attitude indicators and can also be used as visual aids during rescue missions. The paper [3] explains the use of computer vision system to measure spatial position of aircraft flying close to the water surface. Due to their possibility of it versatility as well as easy integration with other systems, they hold an advantage over conventional pitot-static altimeters, and radio altimeters. Radars also have the same potential, however at such low-altitudes they cannot be used due to errors which can be up to 10–20 cm [4]. In this paper, only the task of using computer vision systems as altimeter for all sea states, which the Ekranoplane operates in, is proposed.

Although, 5 m seems to be an altitude where the aircraft can be flown easily under Visible Flight Rules (VFR), an accurate altitude measurement system is required because such aircraft are very sensitive to small changes in altitude,

Fig. 1 Ekranoplane flying close to sea surface

which need to be immediately reported to the pilot. Two systems using 'Computer vision' were proposed to solve this problem.

- Passive stereoscopic camera system (PSCS)
- Active stereoscopic camera system (ASCS)

Other systems like interferometry systems are good for micro-meter ranges in industries to detect inaccuracies in solid surfaces. TOF cameras are presently an expensive technology [5].

2 PSCS for Determining Altitude

The passive camera system works on the principle of photogrammetry. It uses two cameras placed at an optimal distance from each other—usually 6.5 cm. Their function is to obtain stereoscopic images of the sea surface below the flying aircraft. These stereoscopic images are processed and analyzed, to determine the altitude of the aircraft. [6] The block diagram of the altimeter with passive camera system and its operation is shown in Fig. 2.

In this system, the stereo-camera is placed on a stabilized platform so as to obtain near—vertical images of the water surface. The altitude of the camera system above the water surface is determined by the percent overlap of the stereoscopic images. The percent overlap is determined by cross-correlating the processed stereoscopic images. In low light conditions, this system uses artificial external source of light for illuminating the water surface [7].

When water is perfectly smooth, this method could make use of the reflection of the aircraft on the water surface to determine height, however, it would be preferred to use active camera system instead, because reflections may not be very clear and can differ due to the different viewpoints of the cameras. The original drawback of the system is that it cannot work in low light conditions.

Experiments were done using a stationary camera and a uniform stationary object on ground, the maximum error was found to be 0.09 cm for a measured height of 1.5 m. Details of the experiment and results were presented in [6]. Books

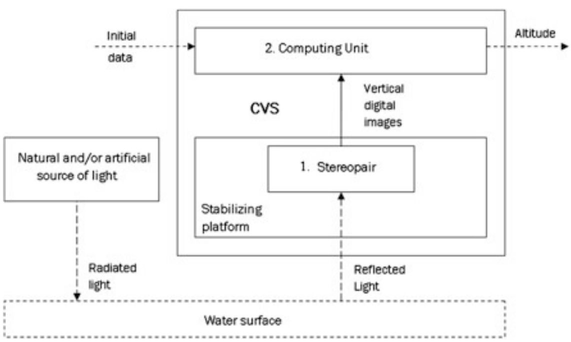


Fig. 2 Functional block diagram of PSCS (translated from [2])

like [5] and [8] also state that the errors using this method are remarkably low when calibrated cameras are used. The method is still to be practically tested on sea using calibrated cameras. However, preliminary experiments prove that the method is much more superior to the existing technologies like radars and radio-altimeter.

3 ASCS for Determining Altitude

The active camera system uses an array of diffused LED lights mounted between the stereoscopic pair of cameras. The diffused LED light is reflected from the surface, and part of the light falls on the objective of the stereoscopic pair of cameras. In this method, image segmentation and analysis is used to detect the reflection of at least one of the LEDs light on a smooth mirror-like water surface. Knowing the orientation parameters of the cameras and their position relative to the light source, it is possible to determine not only altitude, but also the angular orientation of the object with respect to the water surface. Thus it eliminates the need of using a stabilization platform. The block diagram of the altimeter with active camera system and its operation is shown in Fig. 3.

This system works when there is no disturbance on the water surface. Also, it works well on dark waters and low light conditions [7].

4 System Integration

The aim is to design an altimeter based on computer vision that will work in all flight conditions of an Ekranoplane. Let us consider all the conditions under which the active and passive systems work or will work when combined:

- Bright-light and disturbed sea surface: Passive
- Bright-light and mirror-like sea surface: Active
- Low-light and disturbed sea surface: Passive combined
- Low-light and mirror-like sea surface: Active

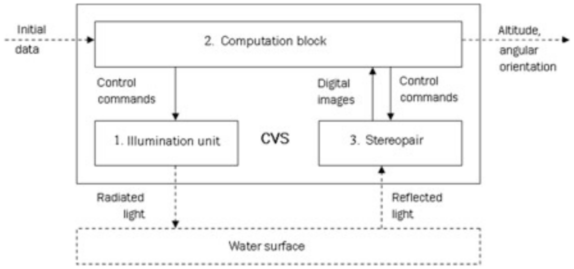


Fig. 3 Functional block diagram of ASCS (translated from [7])

Although, the camera systems can be used independently, there is no guarantee of whether the water surface will remain smooth or in the disturbed condition throughout the duration of flight. Hence, an effective integration of the systems is required. There are two parts to such system integration:

- Hardware integration
- Software integration

5 Hardware Integration

The hardware of the integrated system is very much similar to the active camera system hardware that was proposed in [3]. The only addition to the system is a stabilized platform and Light Emitting Diode (LED) arrays placed at the edges for PSCS. The LEDs in these arrays are closely placed so as to act as an artificial source of light for operating in low light conditions. They are preferred to be white in color. Although in Fig. 4 only a few arrays of white LEDs are shown, some arrays can be added or subtracted according to requirement, space available and maximum altitude of flight for the particular type of Ekranoplane. The arrays of LEDs placed near the center (green in color) are for ASCS system. The LEDs are chosen to be green because green light has higher reflectance than red light, which penetrates through water, but doesn't reflect back to the sensor as much. [7, 9]. They are placed such that reflection of at least one LED is visible in the stereoscopic images. These LEDs are not used for the purpose of illumination, but for the purpose of determining altitude from their reflection on the water surface. Both the systems rely on the same sort of electric circuits and image processors, and only their algorithms differ.

6 Software Integration

PSCS was originally incapable of working during the night. In the integrated system, PSCS can make use of the LED arrays to illuminate the scene. However, the first step would be to determine the low light condition threshold and

Fig. 4 A conceptual view of the integrated system (modified from [3])

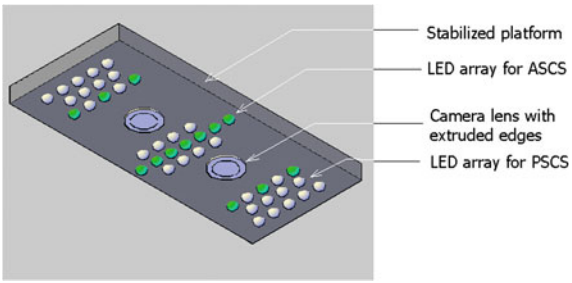


Fig. 5 Decreasing brightness of water surface

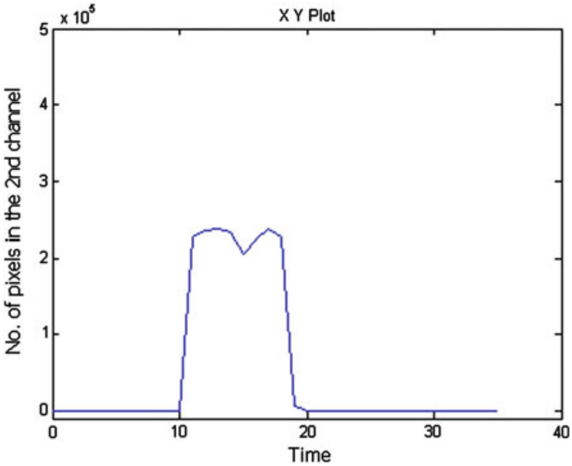


Fig. 6 Graph showing the variation in number of pixels in the second channel with respect to time

accordingly to activate the LEDs. The option of keeping the LED arrays switched on throughout the flight is not a good one, as they may create unnecessary disturbance and gleam for both the PSCS and ASCS systems.

It is easy to determine whether an image is obtained in low light condition or good light condition using the histogram function available in MATLAB. A video of 20 s of varying brightness levels was obtained in an indoor set-up. The brightness drops at 10 s and then increases again 19 s in the video. Three frames of this video with decreasing brightness (extracted from 3, 6 and 9 s correspondingly) are shown in Fig. 5.

A simulink program was used to detect the brightness level. The program was simulated for 35 s. Here the histogram was divided into ten bands, and the output was taken from the second channel—number of pixels that have an intensity value from 25 to 50 (dark pixels) from the 256 intensity values present in the image frame. The output could be taken from any suitable channel.

The resulting graph in Fig. 6 confirms with the video that frames obtained from 10 to 19 s in the video have low brightness level. Thus the sudden rise indicates the time at which the LEDs should be activated and the sudden fall indicates the time at which the LEDs should be de-activated.

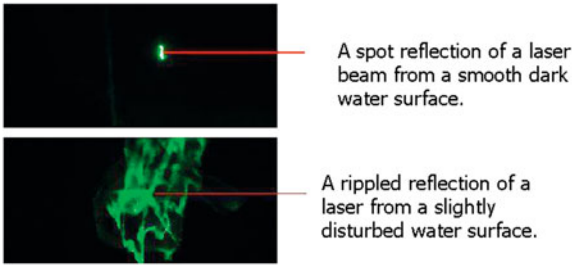


Fig. 7 Types of light reflections depending on the nature of water surface in night conditions

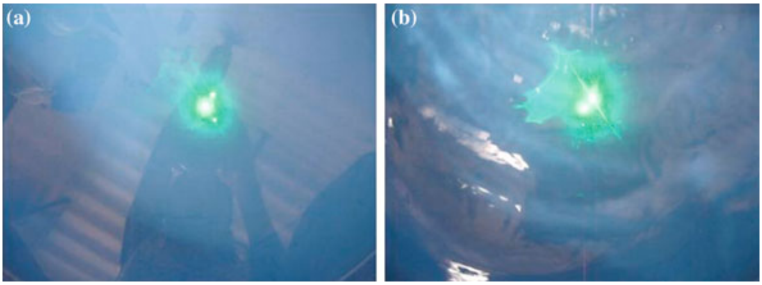


Fig. 8 Still water surface with a laser beam reflection (**a**) and disturbed water surface with a laser beam reflection (**b**)

Now that both the systems have similar capabilities, what remains to be determined is whether the water surface is rough or smooth. This will act as a switch between ASCS activation and PSCS activation. The ASCS LED arrays will be active all throughout the operation of ASCS. This switch can be achieved through the method of image segmentation and then area calculation. Suppose that the present mode of operation is ASCS where the reflection of light on the smooth water surface is segmented according to color. The reflection of light changes according to the nature (texture) of water surface [7]. This is shown in the Fig. 7. If the water surface becomes more disturbed then the light is not at all reflected.

Figure 8 shows two frames from the 35 s video that was obtained with an indoor set-up. In this experiment, the laser beam was allowed to glow throughout the experiment, and the water was disturbed from 8 to 22 s.

The simulink program was used to detect a change from smooth to rough surface as well as rough to smooth surface. A ‘Label’ block determines the number of connected objects in an image frame.

Figure 9 shows the output graph, which demonstrates that when the sea is smooth, the number of connected objects detected is small (below 10) and as the roughness increases the number of connected objects detected have increased. It is evident from it that disturbance causes the light to be reflected from different wave fronts, thus causing the effect of increase in number of smaller connected objects.

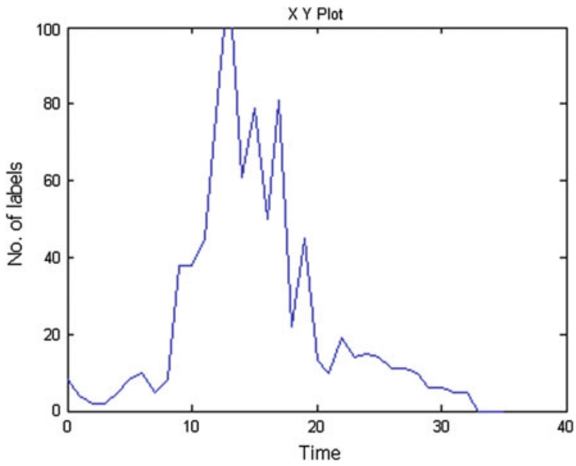


Fig. 9 Graph showing the variation in number of connected objects with respect to time machinery

In case of a rougher sea, the reflection will not be visible at all. In case the number of pixels exceeds a pre-determined value, or is almost absent, then a 'soft switch' is activated, and the operation switches to PSCS. Before completely, switching over, it is necessary to determine the intensity of the scene, and to decide accordingly whether the LED arrays have to be activated or switched off. The ASCS LED array will be switched off once ASCS is deactivated. Only after this, the PSCS starts functioning. Care should be taken in the practical case for such a switching to occur seamlessly, without causing any data loss to the pilot.

The next switch should take place from PSCS to ASCS i.e., when a smooth sea is detected in day light conditions. It is found that when surface of water is smooth, the intensity variation is smooth—presuming that neither the bottom of the sea nor the clouds reflected by the water surface cause any abrupt variations in its intensity. A disturbed water surface will have abrupt variations in intensity levels. Thus it is easy to detect a smooth surface by morphing the image.

Figure 10 shows two frames from the 29 s video that was obtained with an indoor set-up. In this experiment, the water was disturbed from 3 to 24 s. It was performed under day light conditions.

In this simulink program, the image was morphed using 'canny' edge filter and then the 'label' block was used.

Images with a smooth surface will have a long continuous edge as compared to the rough surface image, when the edges break and appear as numerous connected smaller objects. Figure 11 shows the output graph where the number of broken edges increases (above 35) when the water surface is rough (from 3 to 24 s) .

Care should be taken to avoid any ringing effects—i.e., a continuous switch from ASCS to PSCS and vice versa. Wind is the main factor that causes calm spells and disturbed spells on the sea surface. Rapidly changing wind would adversely affect the operation of Ekranoplane, and hence this effect is not a possibility for the present moment.

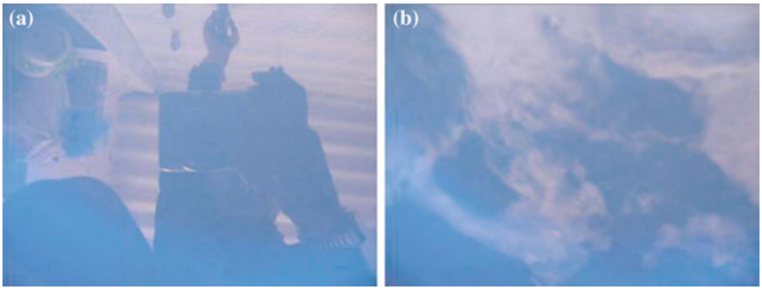
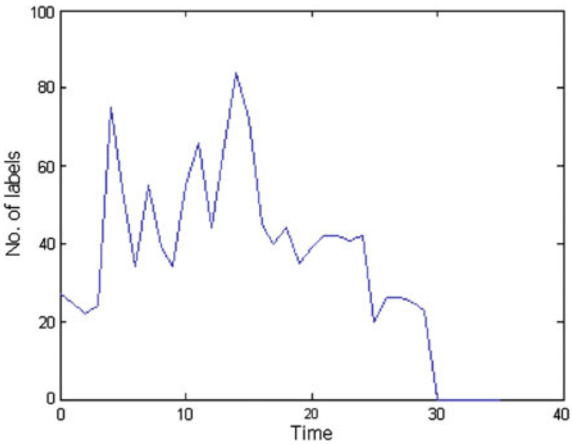


Fig. 10 Still water surface (**a**) and disturbed water surface (**b**) in daylight conditions

Fig. 11 Graph showing the variation in number of connected objects with respect to time machinery



The process of checking the intensity of the scene, as well as checking the nature of water surface are iterative processes and requiring additional computation and time by processor. These checks can be done at certain intervals of time. The values received as part of the checks can be stored for a certain time for the system to record whether light or surface conditions are changing. If no change is recorded, the checks can be performed after a pre-determined increment in time. Another method is to activate the check after the pilot manually chooses between ASCS and PSCS.

7 Conclusion

Thus, through this paper, integration of ASCS and PSCS systems to determine altitude above water surface was proposed so as to work in all light conditions and in all sea states that the Ekranoplane operates in. A method was found to overcome the drawback of PSCS which was originally incapable of working during low light condition. Hardware integration was described in detail using a conceptual view of the system. The software integration is proposed by designing a ‘soft’ switch using different image processing techniques. The efficiency of the underlying algorithms were proven by obtaining videos, processing them through simulink programs and showing the graph results.

References

1. Nebylov AV, Wilson P (2002) Ekranoplane—controlled flight close to the sea. WIT Press, Southapton
2. Liang Y, Alan B, Johnny D (2010) WIG craft and ekranoplan ground effect craft technology. Springer, New York
3. Krysin DY, Nebylov AV (2012) Algorithms of system measuring spatial position of object near water surface. In: XIVth Conference of young scientists navigation and motion control. St. Petersburg (in Russian)
4. Nebylov AV, Jatchevitch G, Nebylov VA (2012) New structures and algorithms of altimeters for controlling vehicles motion at low and superlow altitudes. In: Proceedings on CD, IFAC-EGNCA, Bangalore (in press)
5. Bernd J, Horst H (2007) Computer vision and applications—a guide for students and practitioners. Academic Press, US
6. Rao VP (2011) Measuring flight parameters using digital aerial camera for ekranoplanes. Aerospace young engineer awards, Bangalore
7. Krysin DY, Rao VP (2012) Features of using computer vision for measuring spatial position of object near water surface. Scientific session of SUAI, St. Petersburg (in press)
8. Peter C (2011) Robotics, vision and control, fundamental algorithms in MATLAB. Springer, Berlin
9. Martin S (2004) An introduction to ocean remote sensing. Cambridge University Press, Cambridge

Gateway-Based Modified AODV for MANET and Internet Interconnection

Yushan He, Youngshin Ahn and Jaeho Choi

Abstract This paper describes a new dynamic gateway method which can be used to connect a MANET to the Internet. In the proposed method, every mobile node in a wireless mobile ad-hoc network can configure itself into a gateway when it receives an interconnection request from a neighborhood node while it is in the range of an access point into the Internet. The interconnection mechanism is based on a modified AODV routing protocol that can provide a route between the nodes in the MANET and to the Internet through a gateway. In order to make intra- and inter-communications in and out of the MANET, an extended IP datagram head is also introduced where a tag field is used to distinguish the traffic and also to combine the multiple gateways. The results obtained by computer simulations using the OPNET Modeler show that the proposed dynamic gateway method successfully provides a robust connection between the MANET and the Internet and enables services like e-mail, http, and voices over heterogeneous networks.

Keywords Dynamic gateway · Modified AODV routing · Interconnection

Y. He · Y. Ahn · J. Choi (✉)
Department of Electronic Engineering, CAIIT, Chonbuk National University, Chonju, Chonbuk 561-756, Republic of Korea
e-mail: wave@jbnu.ac.kr

Y. He
e-mail: colinmengyu@hotmail.com

Y. Ahn
e-mail: math1532@nate.com

Mohan S. and S. S. Kumar (eds.), *Proceedings of the Fourth International Conference on Signal and Image Processing 2012 (ICSIP 2012)*, Lecture Notes in Electrical Engineering 222, DOI: 10.1007/978-81-322-1000-9_17,

1 Introduction

A mobile ad hoc network (MANET) is an autonomous network that functions without any established infrastructure or centralized administration and is comprised of free roaming nodes which communicate wirelessly using radio transmissions. Ad-hoc networks can be formed quickly on demand, and are of use in emergency response and to military organizations [1, 2]. An extended ad-hoc network uses multihop forwarding to deliver data. The drawback of pure ad-hoc networks is that they cannot be connected to the hierarchy of the Internet structure due to their routing approach. This interconnection is achieved by using a gateway, which establishes the connection between the MANET and the Internet [3].

In this paper, we show how to provide ubiquitous Internet connectivity for mobile nodes of ad-hoc networks. To achieve this goal, Internet gateways must be discovered by the mobile nodes of the MANET using modified ad-hoc routing protocols. For this purpose, a modified AODV protocol and dynamic gateway concepts are proposed. The proposed method also uses the mobile node as a gateway and utilizes an email application to verify the correctness and feasibility of the approach.

The remainder of this paper is organized as follows. Section 2 describes some basic concepts and related works. Section 3 contains the extended AODV protocol and the description on the dynamic gateway. Section 4 presents the simulation environment and the performance evaluation using network simulations. Finally, the conclusion is made in Sect. 5.

2 Related Works

In this section, the basic AODV concept existing gateway discovery schemes are discussed. AODV is designed for ad hoc mobile networks and of both routing, which establish routes between different nodes that are needed by source nodes. There are three messages which are defined by AODV: Route Errors (RERRs), Route Requests (RREQs), and Route Replies (RREPs) [4]. These messages are used for discovering and maintaining the routes in the network. When a route must be created at the destination, a node broadcasts a route request (RREQ) message and sets a timer to wait for the route reply (RREP) [5].

Several techniques have been proposed for interconnecting MANET with the Internet. However, all existing approaches consider only fixed gateways to connect MANET nodes to the wired Internet [6]. Lie and Parkins [7] proposed mechanisms to construct ad hoc networks and to provide Internet access to MANET nodes. Their approach makes use of a routing protocol within an ad hoc network and a modified Routing Information Protocol (RIP) to interconnect the MANET with the Internet. Broch et al. [8] proposed a solution to the integration of MANET with Mobile IP. They introduced the notion of a border router, which has two interfaces. One interface is connected to the router, which also has two interfaces. The other interface, which is connected to the Internet, is configured so that normal IP routing mechanisms can be used when packets come in or out of the MANET. The interface connected to the MANET uses the dynamic source routing (DSR) protocol to route packets within the MANET. In [9], Ratanchandani and Kravets propose a hybrid scheme using Mobile IP to provide global Internet connectivity to MANET nodes. They used techniques such as TTL scoping of agent advertisements, eavesdropping, and caching agent advertisements. This hybrid approach benefits from the combined use of reactive and proactive approaches.

3 Proposed Method

In the proposed method, we define a particular circumstance: whole nodes have a global IP address and every node has one interface to the MANET. When a normal MANET node moves into AP communication range, it becomes a gateway to connect the MANET and the Internet. The gateway will periodically broadcast an advertisement message (GWADV). Upon receipt of the advertisement, the mobile nodes that do not have a route to the gateway create a route entry for it in their routing tables using the extended AODV protocol. Mobile nodes that already have a route to the gateway update their route entry for the gateway. When a link fails, AODV local repair is triggered. If multiple gateways exist, path options are compared and loaded to the chosen gateway. Our proposed approach allows the mobile node to work as a gateway, as shown in Fig. 1.

3.1 Dynamic Gateway Discovery

As stated previously, the gateway periodically broadcasts a gateway advertisement message, which is transmitted after expiration of the gateway's timer. All mobile nodes residing in the gateway's transmission range receive the advertisement.

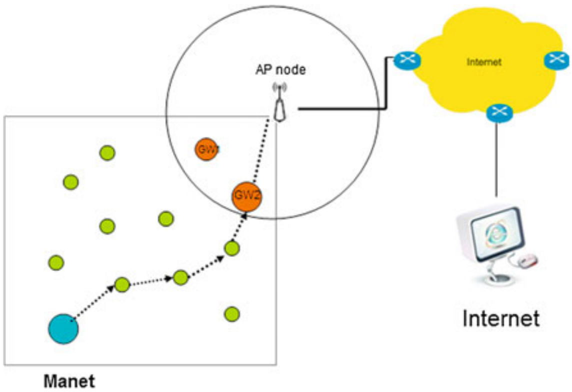


Fig. 1 Interconnection between MANET and Internet

Fig. 2 The gateway advertisement message format

Type	Reserved
GWADV ID	
Sequence number	
Originator address	
Life time	

When a mobile node receives a GWADV, it first checks to determine whether a GWADV with the same original IP address and GWADV ID have already been received during the last BCAST_ID_SAVE seconds. If such a GWADV message has not been received, the message is rebroadcast. Otherwise, if a GWADV message has been received, the newly received GWADV is discarded. Hence, duplicated GWADVs are not forwarded and the advertisement is flooded through the whole network without causing excessive congestion. Figure 2 illustrates the GWADV message format.

3.2 Path Establishment

In the path establishment phases, the new algorithm is based on the RREQ/RREP of the AODV protocol. The modification consists of a special flag which is called the I-flag. The flag includes two values: 1 and 0. I_flag = 1 indicates that the route request requires a gateway connection. This is equal to zero standing for an ordinary route request. Figures 3 and 4 show the format of the RREQ_I and RREP_I messages. Assume that a mobile node S wants to communicate with another node D and that S does not have any route to D in its routing table. Using extended AODV as the ad hoc routing protocol, S broadcasts a RREQ_I, requesting a route to D.

When an intermediate mobile node receives a RREQ_I message, it searches its routing table for a route to the destination and generates a reverse route table. If D is a mobile node, the I_flag value is equal to zero in the RREQ_I message. The node D will unicast back a RREP_I (I flag = 0) to generate a forward route table

Fig. 3 The RREQ_I message format

Type	J	R	G	D	U	I	Reserved	Hop count
RREQ ID								
Destination local address (GW MANET address)								
Destination sequence number								
Originator local address								
Originator sequence number								

Fig. 4 The RREP_I message format

Type	R	A	I	Reserved	Prefix size	Hop count
Destination local address						
Destination sequence number						
Originator local address						
Life time						

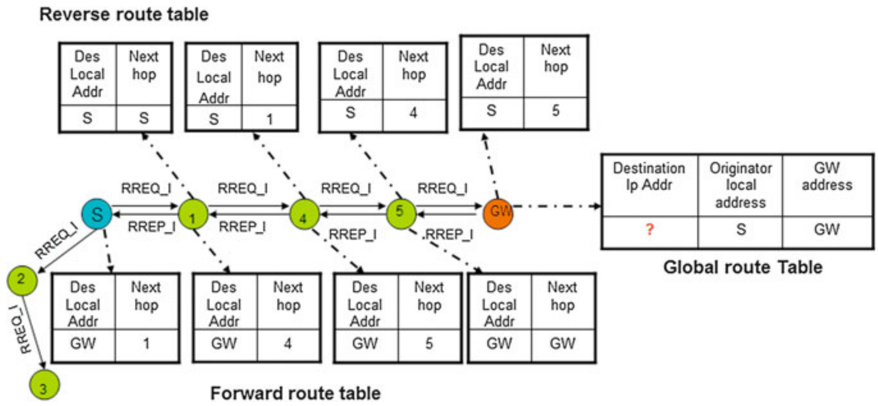


Fig. 5 The modified AODV in MANET

Fig. 6 Extended IP datagram header format

Version	IHL	Type of service	Total length
Identification			Fragment offset
Time to live		Protocol	Header checksum
Source address (4 bytes)			
Destination address			
GW MANET address			

for S. However, if D is a fixed node (located on the Internet), the intermediate node must not send a RREP_I back to the originator of the request message even if the route is found. If the intermediate node sends a RREP_I back to the originator of the RREQ_I message, the originator thinks that the destination is a mobile node that can be reached via the intermediate node. When a gateway receives a RREQ_I (I_flag = 1), it looks in its routing table to search for the originator local address and RREQ ID specified in the RREQ_I message. If the address is not found in the routing table, the gateway has to send a RREP_I (I_flag = 1) back to the originator of the RREQ_I. The gateway then generates a global route table. Figure 5 shows the entire process.

In general, it is necessary for the source node to create a route entry for the fixed node in its routing table, there is a disadvantage. If a mobile node desires to communicate with many fixed nodes, its routing table will grow rapidly. In this paper, we proposed a method that extends the IP datagram to solve this problem. Figure 6 shows the extended IP datagram format.

If a mobile node sends data to many internet nodes, it uses the same route path, which will rapidly decrease route overhead. The global route table will be complete through the extended IP datagram. If node S delivers data to Internet node D, the entire process proceeds as shown in Fig. 7.

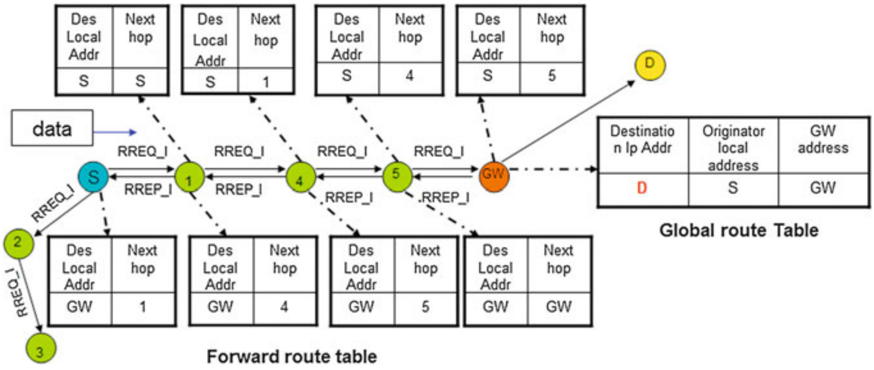


Fig. 7 The data exchange between MANET and Internet by the gateway node

Table 1 Common parameters for simulations

Parameters	Value	Parameters	Value
Network size	1600 m^2	Mobile speed	2–5 m/s
Range	100 m	Physical characteristics	Direct sequence
Simulation time	100 s	Data rate	1 Mbps

4 Performance Evaluation

The proposed strategy is implemented using the OPNET simulator [10]. We will demonstrate the simulation results of the proposed dynamic gateway approach with a MANET interconnect with the Internet and compare the results to those of our previous method.

In order to prove that our proposed method is feasible and correct, we demonstrate its use with four scenarios: an e-mail service scenario, a gateway change scenario, a gateway choice scenario, and a comparison scenario.

(1) **Email service scenario**: simulates a process in which Email is received and replied to by using a dynamic gateway.
(2) **Gateway change scenario**: shows the process by which a new gateway replaces the previous gateway as the old gateway moves out of AP communication range.
(3) **Multiple gateways scenario**: shows the process of selecting between multiple gateway choices by using a combination algorithm.
(4) **Comparison scenario**: performances in terms of throughput and delay are compared with those of the conventional method. The simulation results show that our method is superior to the previously described method.

This section describes various simulation configurations. The simulation model and initial locations (X and Y coordinates) of the nodes are obtained using a

Table 2 Simulation parameters for scenario 2 and 3

Parameter	Value	Parameter	Value
Mobile node number	6	Gateway number	2
Hop counts	3	Packet size (bytes)	1600
Rate requests/hour	1800	Type of service	Best effort

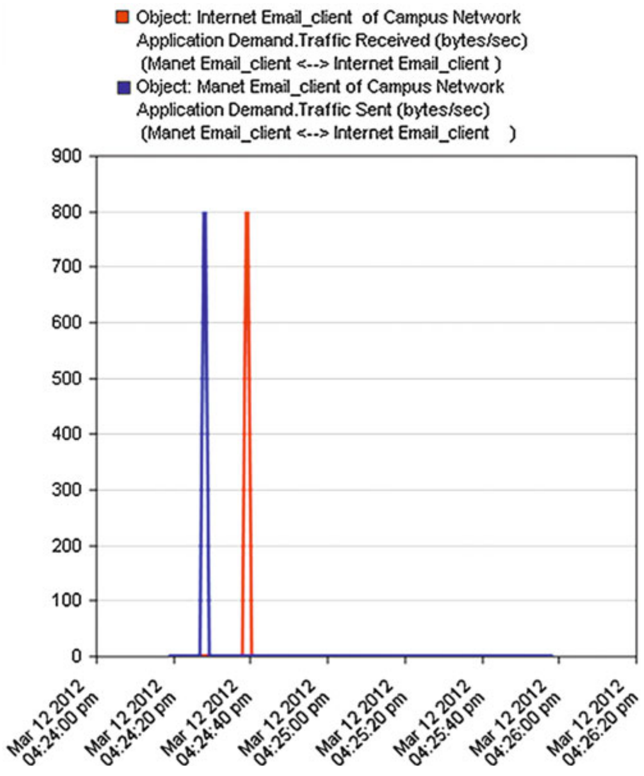


Fig. 8 A MANET node sends an email to an Internet node

uniform distribution. Table 1 shows common simulation parameters and configurations. The parameters of the MANET nodes are shown in Table 2.

(1) **Email service scenario result**. Figures 8 and 9 show the process of the MANET node sending an 800 byte “Email” to the Internet node the Internet node receiving it about 10 s later. Figure 10 illustrates why the Internet Client received the “Email” in 10 s. The first point indicates the MANET Client AODV discovery time and the second point is packet delivery time. When the values from the two figures are added together, the number is 10 s. If there are more hop counts in the MANET, a longer discovery time may be needed to receive the “Email”. Similarly, the delivery time increases as packet size

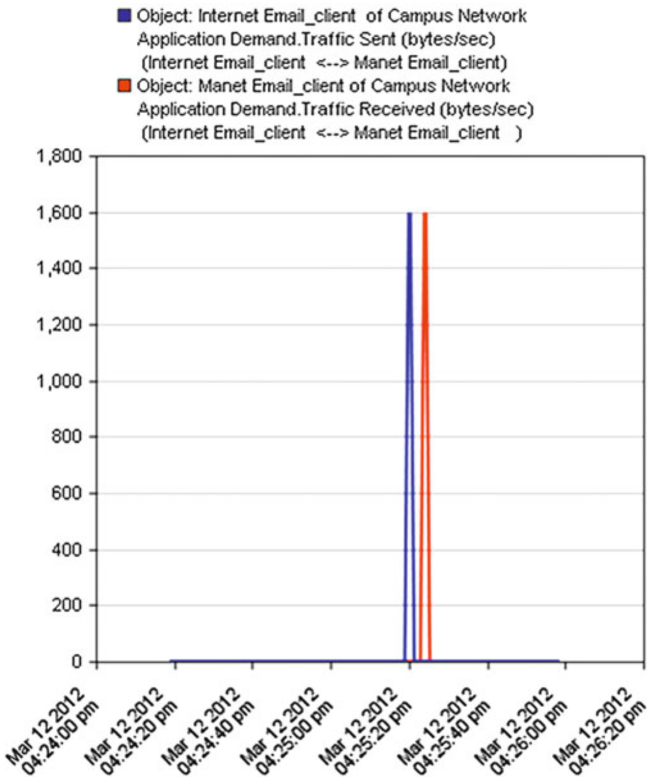
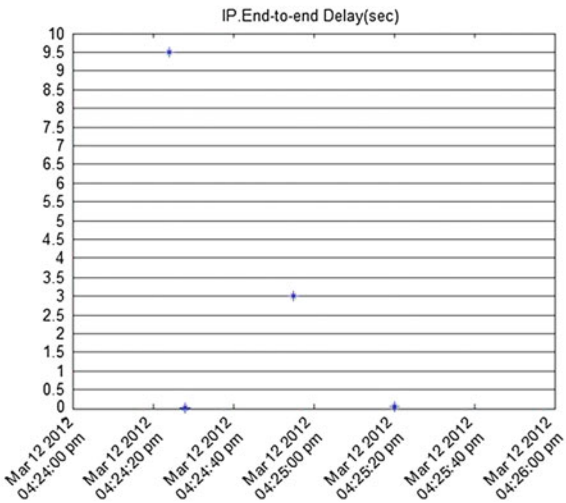


Fig. 9 The Internet node replies this email to MANET node

Fig. 10 End-to-end delay for sending and receiving



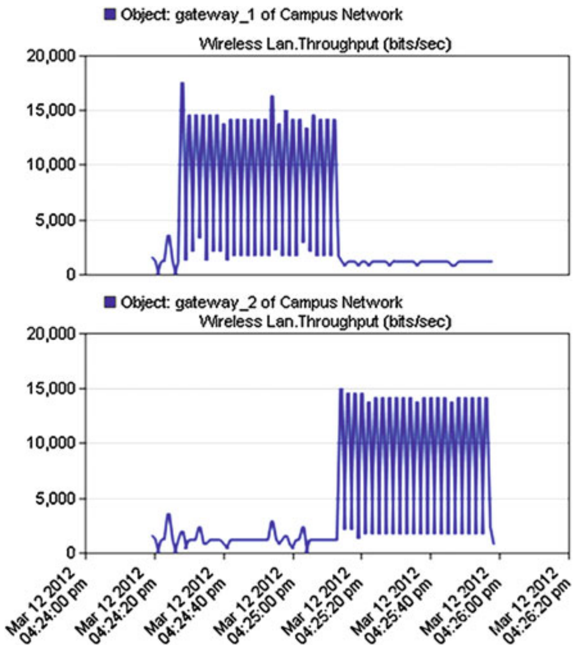


Fig. 11 Gateway changes from G1 to G2 in a gateway change scenario

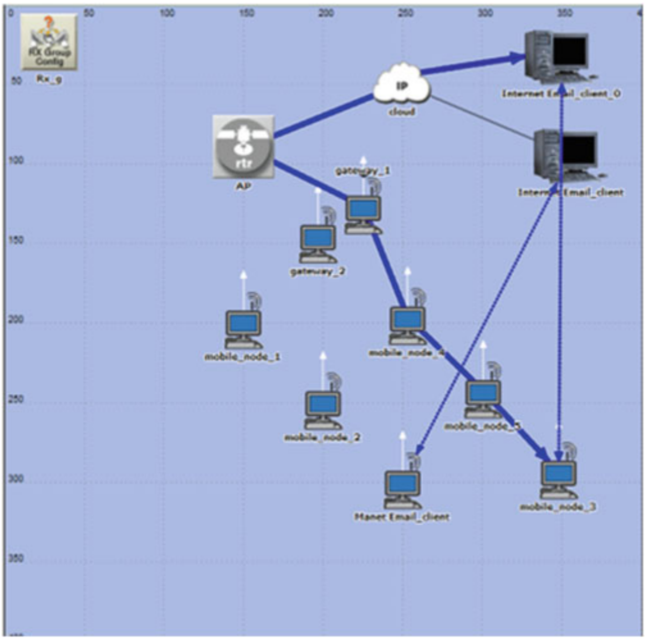


Fig. 12 Gateway_1 path display in the multiple gateway scenario

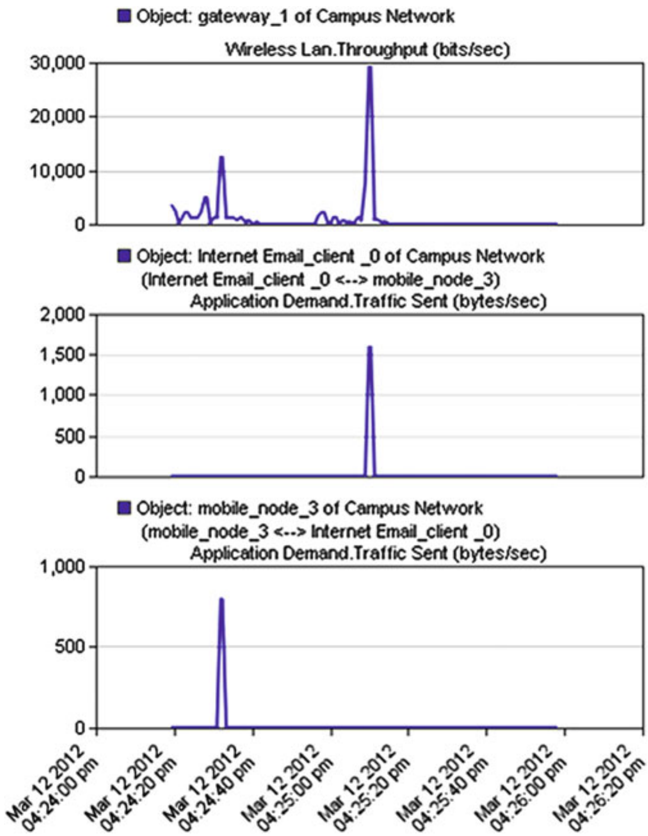


Fig. 13 Gateway_1 throughput in the multiple gateway scenario

increases. Figure 10 shows the Internet Client replying to the "Email" situation. In Fig. 10, the third point indicates the discovery time when the gateway is moving. However, these gateways still store information about the path. Comparing the first and third points, the latter evidently decreased in the former. Moreover, the fourth point is slightly higher than the second point because the reply email size is bigger than the original email size.

(2) **Gateway change scenario result**. Figure 11 illustrates the whole process result. Initially, Gateway_1 acts as a bridge connecting MANET and the Internet, and the MANET Client sends a data packet to the Internet Client every 2 s. At 25 min 20 s, Gateway_1 moves out of AP communication range and becomes a normal mobile node, which leads to communications being interrupted. Under the AODV protocol, when the Gateway_1 upstream node detects a break, it does not send a RERR message back to the originator of the route, but instead tries to repair the route while buffering data packets. The upstream node sends a RREQ_I (I_flag = 0) to the destination of the original route. Gateway_2 receives the RREQ_I message and transmits a RREP_I

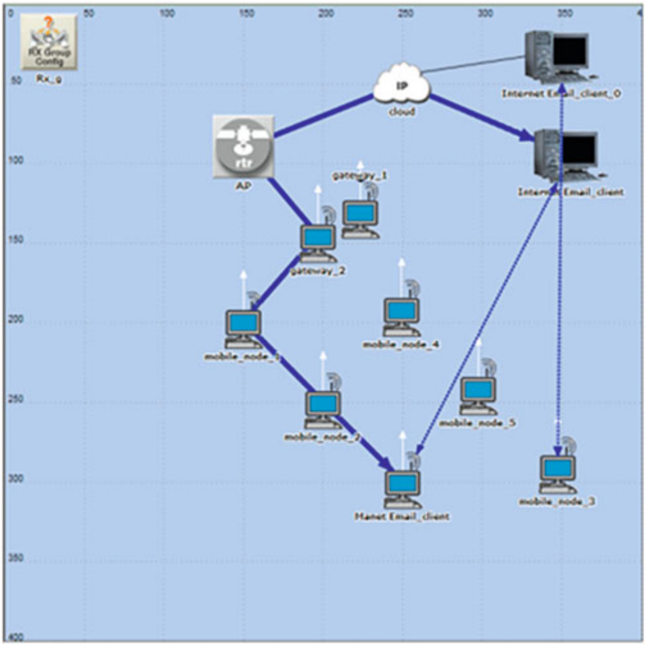


Fig. 14 Gateway_2 path display in the multiple gateway scenarios

message to the upstream node. Gateway_2 is used instead of Gateway_1 and path repair is finished.

(3) **Multiple gateway scenario result**. The routes to both Internet gateways, GW1 and GW2, have the same hop count. Figure 12 shows mobile_node_3 communicating with Internet Email_Client_0 through gateway_1. The path from mobile_ node_3 to gateway_1 is depicted by solid arrows. From the gateway throughput perspective, mobile_node_3 sends a data packet to Internet Email_Client_0 in 24 min 40 s, and Internet Email_Client_0 replies with a data packet to mobile_node_3 in 25 min 10 s. Gateway_1 throughput is significantly increased by a factor of two, as shown in Fig. 13. In Mobile_-node_3 communications, the MANET Email_Client is also communicating with the Internet Email_Client. In this case, the MANET Client chooses another path which interconnects with the Internet via gateway_2. This selection avoids path congestion and resulting data loss. The gateway_2 choice result is shown in Figs. 14 and 15.

(4) **Comparison scenario result**. In this scenario, we compare our proposal with the DSR method proposed in [11]. The parameters are the same as in [11]. As shown in Fig. 16, the average delay from the source to the destination in our proposed method is less than in the method described in [11] because all of the routing information for the path is taken from the source to the destination in the DSR. With increasing hop counts, route overload and delay will increase.

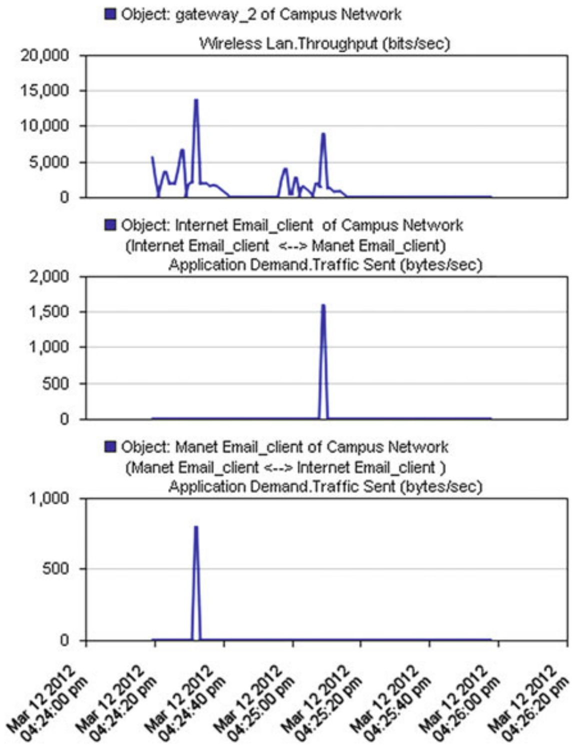


Fig. 15 Gateway_2 throughput in the multiple gateway scenario

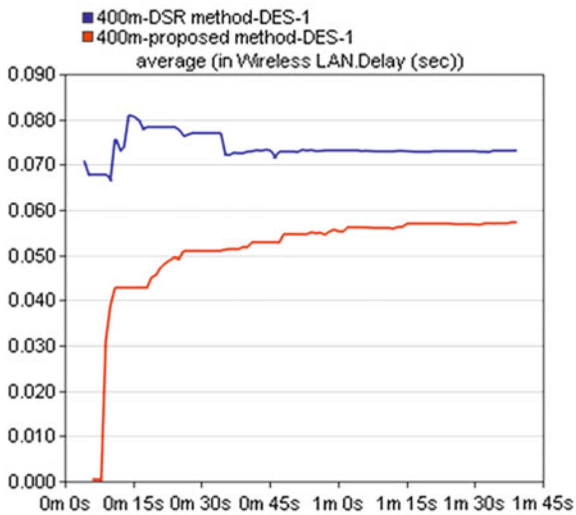
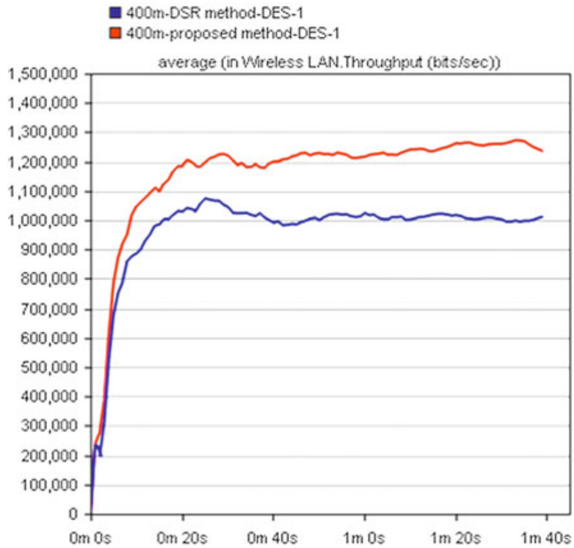


Fig. 16 The average delay from a source to a destination in the comparison scenario

Fig. 17 The throughput of the source node in the comparisons scenario



This reduction in the delay improves the overall performance of system. Hence the throughput of the system is higher than using a conventional method. Because the source node sends data packets on a less congested path, the data sending rate is maintained as shown in Fig. 17.

5 Conclusion

In this paper, we have presented a dynamic gateway, which acts as an interface between MANET and the Internet. The ad hoc routing protocol AODV has been extended to route packets, not only within a MANET but also between a wireless MANET and the wired Internet. In this paper, four scenarios for evaluating dynamic gateways were presented. To demonstrate feasibility of the proposed dynamic gateway, the email service is tested in the first scenario. The result shows that email traffic supported robustly end-to-end. The second scenarios illustrate how the gateways can change their position and are replaced. The proposed method allows for smooth transitions from the old gateway to the new gateway. When faced with emergencies, the proposed method can demonstrate a superior reliability. The last scenario presents gateway selection from multiple gateways. The result shows that the proposed method can be adapted to complex environments. For the future work, the proposed dynamic gateways will be extended to support multimedia streaming.

References

1. Hamidian A (2003) A study of internet connectivity for mobile Ad Hoc networks in NS2. Department of Communication Systems, Lund Institute of Technology, Lund University, Sweden
2. Ohta T, Inoue S, Kakuda Y (2007) A hierarchical multicast routing based on inter-cluster group mesh for mobile ad hoc networks. LNCS 4864:32–43
3. Rosenschon M (2007) Internet gateway discovery for mobile Ad Hoc networks. School of Engineering and Mathematical Sciences, CITY University London, London
4. Perkins C, Belding-Royer C, Das S (1999) Ad Hoc on-demand distance vector (AODV) routing. IETE Internet Draft, pp 90–100
5. Ma X, Sun B (2009) Design and implementation of the MANET and internet interconnection based on AODV. Intel. Info. Tech. application workshop, pp 108–111
6. Sun Y, Belding-Royer EM, Perkin CE (2002) Internet connectivity for Ad hoc mobile networks. Int J Wirel Inf Netw 9 (special issue on mobile Ad Hoc networks: standards, research, applications)
7. Lei H (1997) Parkins: Ad Hoc networking with Mobile IP. In: Proceedings of the 2nd European personal mobile communication Conference, pp 75–82
8. Broch J, Maltz D, Johnsonk D (1999) Supporting hierarchy and heterogeneous interfaces in multi-hop wireless Ad Hoc networks. In: Proceedings of the workshop on mobile computing, pp 370–375
9. Ratanchandani P, Kravets R (2003) A hybrid approach to Internet connectivity for mobile Ad Hoc networks. In: Proceedings of IEEE wireless communications and networking conference, pp 16–20
10. Information on http://www.opnet.com
11. Baniya B, Wang X, Heo U, Choi J (2010) Internet gateway routing for MANET. In: Proceedings of cyber-enabled distributed computing and knowledge discovery, pp 199–204

Multiple Classification Method for Analysis of Liver Lesion with Focal Liver Segmentation Techniques for CT Image

H. N. Suma, Appaji M. Abhishek, M. Chaithanya Lakshmi and Y. Veena

Abstract Liver is the biggest organ in the human body, which has basic functionalities like storage of glucose, producing bile juice. In this paper the analysis of focal liver lesions of CT image using multiple classifier methods is addressed. The analysis scheme includes pre-processing, segmentation, feature extraction and classification. Pre-processing of the image is through anisotropic diffusion to reduce noise and inhomogenities in the image. Segmentation with seeded region growing is followed by binary masking to extract liver from CT image. Five features that are based on texture properties such as mean, standard deviation, entropy, root mean square (rms) value and energy are extracted. Classification using Neural Networks, K-Nearest Neighbour (KNN) and Support Vector Machine (SVM) to differentiate the normal image and abnormal imageis performed. In Neural Networks receiver operating curve and confusion matrix are used to evaluate the performance of the system for different number of hidden layers. In KNN and SVM the performance of the system is evaluated based on sensitivity, specificity, accuracy and predictive positive value. Finally, comparison of multiple classification methods is done to evaluate the best classification approach for focal liver lesions of CT image.

Keywords Anisotropic diffusion · Cavernous hemangioma · Hematoma · K-nearest neighbor · Livercyst · Neural networks · Support vector machine

H. N. Suma (✉) · A. M. Abhishek · M. Chaithanya Lakshmi · Y. Veena
Department of Medical Electronics, BMS College of Engineering, Bangalore, India

Mohan S. and S. S. Kumar (eds.), *Proceedings of the Fourth International Conference on Signal and Image Processing 2012 (ICSIP 2012)*, Lecture Notes in Electrical Engineering 222,
DOI: 10.1007/978-81-322-1000-9_18,

1 Introduction

The liver is the largest gland in the body, which has basic functionalities like storage of glucose, producing bile juice. It is situated in the upper part of the abdominal cavity occupying the greater part of the right hypochondriac region, part of the epigastric region and extending into the left hypochondriac region.

Basic liver diseases areLiver cyst, Hematoma, Cavernous hemangioma [1]. The features of each type are considered for the analysis and classification. The gross features are:

1.1 Liver Cyst

The contour of the liver cyst is smooth, but the contrast between the normal liver tissue and the liver cyst tissue is high. In other words, the gray level of the liver cyst tissue is much darker than the normal liver tissue. After the contrast media injection, the gray level of the normal liver tissue would be drastically enhanced, and so does the contrast.

1.2 Hematoma

Most of the gray levels of the hepatoma tissues are darker than the normal liver tissue. However, some gray levels in hepatoma tissues might be similar to parts of the normal liver tissues, and the contours of the hepatoma tissues vary from case by case. Thus, it is not easy to discriminate the hepatoma tissues from the normal tissues by the properties of the gray levels or the contours of the suspected regions. The surface of the hepatoma tissues is coarser than that of normal tissues. Thus, taking the contrast media injection into the blood vessels can make the gray levels of the hepatoma tissues more clear to identify.

1.3 Cavernous Hemangioma

Cavernous hemangioma is benign hepatic masses. The hemangioma is composed of large thin-walled blood vessels, lined with flattened epithelium and separated by fibrous spaces filled with venous blood. Thus, the contours of the hemangioma are close to a circle in the CT images, and the gray levels are darker than the normal liver tissues. After injecting the contrast media, the gray levels of the hemangioma tissue can be slightly enhanced.

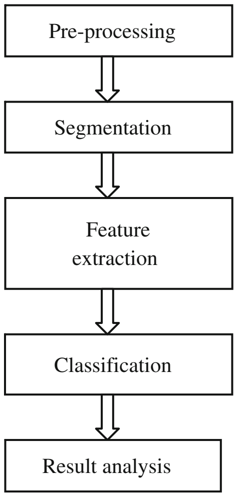


Fig. 1 Methodology

The basic methodology (Fig. 1) comprises of different components, each contributing to the final classification. The proposed methodology consists of pre-processing, segmentation, feature extraction, classification. Pre-processing enhances the contrast of the image, reduces the noise and inhomogeneities in the image and makes the image isotropic. Segmentation will be performed to obtain regions of interest. Texture features appropriate for the classification of specified diseases are will be extracted. Classification based on the textures features is done [2]. Comparison of multiple classification methods is done to evaluate the best classification approach for focal liver lesions.

2 Anisotropic Diffusion

The images contain of the image itself and noise in it. Where noise can be defined as random, little disturbances of the image. Here anisotropic diffusion filtering method is used to reduce the noise. Since in the further steps the segmentation technique is used the noise must be reduced in the image to improve the segmentation of the image. Anisotropic filtering eliminates aliasing effects. Anisotropic diffusion preserves detail at extreme viewing angles in the image.

Anisotropic diffusion is currently one of the most powerful noise reduction techniques in the field of computer vision.This technique takes into account the local structures found in the image (Fig. 2) to filter noise, preserve edges and enhance some features (Fig. 3). Pioneered in 1990 by Perona and Malik, anisotropic diffusion is also called Perona-Malik equation [3]. The method is based on the numerical solution of nonlinear partial differential equation on two dimensions image. The method is then extended into three dimensions. In this

Fig. 2 Original image

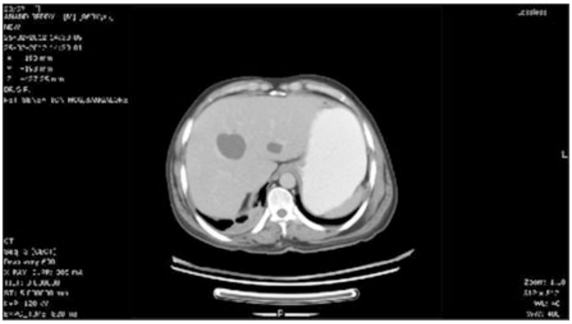


Fig. 3 Filtered image

method, smoothing is modeled as a diffusion that is allowed along homogeneous regions and inhibited by region boundaries. The partial differential equation used to describe the diffusion process is the following:

$$\frac{\partial}{\partial t} I(v,t) = div\,((v,t)\nabla I(v,t)) \tag{1}$$

I(v, t) = div(c(v, t)∇I(v, t)) (1) where div is a divergence operator and ∇ is a gradient operator. I in our case is the 3D volume of CT image and v = (z, y, x) is the coordinate vector. Writing f = (fx, fy, fz), the divergence of vector-valued function f is given by the formula:

$$div f = \frac{\partial fx}{\partial x} + \frac{\partial fy}{\partial y} + \frac{\partial fz}{\partial z} \tag{2}$$

And the gradient of a scalar function f is defined as:

$$\nabla f = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z} \right) \tag{3}$$

Finally, the divergence of ∇f can be written as:

$$div\,(\nabla f) = \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial x}\right) + \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial y}\right) + \frac{\partial}{\partial z}\left(\frac{\partial f}{\partial z}\right) \tag{4}$$

At each voxel, the diffusion strength is controlled by the so-called diffusion coefficient c(v, t) with t is the process ordering parameter used to enumerate iteration steps. The diffusion coefficient c(v, t) depends on the image gradient magnitude ∇I. It should decrease where the gradient magnitude increases so that image regions of high contrast undergo less diffusion, whereas uniform regions are diffused with the same intensity in all directions. Hence the edges can be preserved while removing noise from the image. Two different diffusion functions have been suggested in

$$C1(v,t) = \exp\left(-\left(\frac{|\nabla I(v,t)|}{k}\right)^2\right) \tag{5}$$

$$C2(v,t) = \frac{1}{1 + \left(\frac{|\nabla I(v,t)|}{k}\right)^2} \tag{6}$$

In this work, $c1$ (v, t) is used. This is reported in [3], where $c1$ possesses a much stronger edge enhancing capacity than $c2$. The parameter κ controls the sensitivity to edges and is chosen as a function of the noise in the image.

From the above performance Fig. 3, filtered image is obtained, were the improvements can be seen as preservation of the edges, only smoothing between edges.

3 Segmentation

Segmentation distinguishes objects from background that is it divides as image into parts that have a strong correlation with objects or areas of the real world contained in the image [4]. Simply partitioning of an image into several constituent components is called “Segmentation”. Another word for object detection is Segmentation. The object to be segmented differs greatly in contrast from the background image.

In the present paper, seeded region growing technique has been used. This is followed by binary masking technique to extract the liver from the grown region, in the region growing step. In the region growing the seed point is selected manually and the required region will be grown in the image (Fig. 4). Presently the seed point is selected manually and the region is grown

Binary masking: Binary is used to change specific bits in the original values to the desired setting(s) or to create a specific output value (Fig. 5). A binary mask is used to change one or more bits from 1 to 0 or vice versa using Boolean operation.

Fig. 4 Input image for the region growing

Fig. 5 Region grown Image

The input image in Fig. 6 is read, followed by detection of entire cell in the image, dilation of the image, filling of gaps in the interior of the image, removal of concentrated objects on border and finally smoothening of the object to be extracted as shown in the Fig. 7.

4 Feature Extraction

In the present study features are extracted based on texture base : Mean, Standard deviation, RMS value, Energy and Entropy [5].

Fig. 6 Input image for the binary masking

Fig. 7 Extracted liver

4.1 Mean

The mean is the mathematical average of set of pixels. The average is calculated by adding up two or more pixel values and dividing the total by the number of pixels.

$$Grayavg = \frac{1}{N}\Sigma_{(x,y)\in ROI}g(x,y) \tag{7}$$

where g(x, y) denotes the gray level at pixel (x, y), and N is the total number of pixel inside the ROI.

4.2 Standard deviation

This feature is represented by the symbol σwhichshows how much variation or "dispersion" exists from the average (mean, or expected value). A low standard deviation indicates that the data points tend to be very close to the mean, whereas high standard deviation indicates that the data points are spread out over a large range of values.

$$\sigma = \sqrt{\frac{1}{N}\Sigma_{i=1}^{N}(xi-\mu)^2 \text{where}\, \mu = \frac{1}{N}\Sigma_{i=1}^{N}xi} \tag{8}$$

The result is the square root of an unbiased estimator of the variance of the population from which X is drawn, as long as the image consists of independent, identically distributed samples.

4.3 RMS Value

The root mean square (abbreviated RMS or rms), also known as the quadratic mean, is a statistical measure of the magnitude of a varying quantity. It is especially useful when varieties are positive and negative

$$Xrms = \sqrt{\frac{1}{2}(x1^2 + x2^2 + \cdots + xn^2} \tag{9}$$

4.4 Energy

Energy is nothing but intensity per unit area.

$$E(\theta, d) = \Sigma_{i=0}^{L-1}.\Sigma_{j=0}^{L-1}\, G(\theta, d; i; j).^2 \tag{10}$$

4.5 Entropy

E = entropy(I) returns E, a scalar value representing the entropy of grayscale image I. Entropy is a statistical measure of randomness that can be used to characterize the texture of the input image. Entropy is defined as

$$En = -\Sigma_{i=0}^{G-1}P(i) * \log P(i) \tag{11}$$

where, p contains the histogram counts returned from imhist. By default, entropy uses two bins for logical arrays and 256 bins for uint8, uint16, or double arrays. i can be a multidimensional image. If i has more than two dimensions, the entropy function treats it as a multidimensional grayscale image and not as an RGB image.

5 Classifications

The texture features extracted are used to classify the CT images of liver into normal liver and diseased liver using three different classifiers (Neural networks, K-NN and SVM). The classifiers use a set of 36 images, containing 5 features each. This is applied to the network in form of [5 × 36] matrix, with the rows representing the number of features and columns representing the images. The network is also provided with a target input vector. This target input is a [1 × 36] matrix. Each element of the matrix is the desired output. The prior knowledge is very important because the weights are adjusted according to the desired outputs provided in the target input vector [6].

The testing data consists of 12 images, containing 5 features each. This is applied to the network in form of [5 × 12] matrix, with the rows representing the number of features and columns representing the images. This data has not been trained to classify.

Then comparison of performance of each classifier is done to evaluate the best classification approach for focal liver lesions.

5.1 Neural Network Classifier

The Neural Network classifier used in this study is based on feed-forward back—propagationmethod [7–9]. A feed-forward neural network is an artificial neural network where connections between the units do not form a directed cycle. This is different from recurrent neural networks. In this type of network, the information moves in only one direction, forward, from the input nodes, through the hidden layers (if any) to the output nodes. There are no cycles or loops in the network. In this study the neural networks classification was done using MatlabNeural Networks Toolbox [10], which is a set of tools that include GUIs, wizards and functions that allow any user to use and experiment with neural networks with minimal effort. The neural network training and testing process is carried out for one, two, eight, five and four hidden layers.

5.2 K-Nearest Neighbor Classifier

Instance-based classifiers such as the KNN classifier operate on the premises that classification of unknown instances can be done by relating the unknown to the known according to some distance/similarity function. The intuition is that two instances far apart in the instance space defined by the appropriate distance function are less likely than two closely situated instances to belong to the same class.

Unlike many artificial learners, instance-based learners do not abstract any information from the training data during the learning phase. Learning is merely a question of encapsulating the training data. The process of generalization is postponed until it is absolutely unavoidable, that is, at the time of classification. This property has lead to the referring to instance-based learners as lazy learners, whereas classifiers such as feed-forward neural networks, where proper abstraction is done during the learning phase, often are entitled eager learners.

Classification (generalization) using an instance-based classifier can be a simple matter of locating the nearest neighbour in instance space and labeling the unknown instance with the same class label as that of the located (known) neighbour. This approach is often referred to as a nearest neighbour classifier.

5.3 SVM Classifier

Support vector machine (SVM) is a computer algorithm that learns by example to assign labels to objects. It is a concept in statistics for a set of related supervised learning methods that analyze data and recognize patterns, used for classification and regression analysis. The standard SVM takes a set of input data and predicts, for each given input, which of two possible classes the input belongs to, making the SVM a non-probabilistic binary linear classifier. An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall on.

More formally, a support vector machine constructs a hyperplane or set of hyperplanes in a high- or infinite-dimensional space, which can be used for classification, regression, or other tasks. Intuitively, a good separation is achieved by the hyperplane that has the largest distance to the nearest training data point of any class (so-called functional margin), since in general the larger the margin the lower the generalization error of the classifier.

Whereas the original problem may be stated in a finite dimensional space, it often happens that the sets to discriminate are not linearly separable in that space. For this reason, it was proposed that the original finite-dimensional space be mapped into a much higher-dimensional space, presumably making the separation easier in that space. To keep the computational load reasonable, the mappings used by SVM schemes are designed to ensure that dot products may be computed easily in terms of the variables in the original space, by defining them in terms of a kernel function K(x,y) selected to suit the problem.

The hyperplanes in the higher-dimensional space are defined as the set of points whose inner product with a vector in that space is constant. The vectors defining the hyperplanes can be chosen to be linear combinations with parameters α_i of images of feature vectors that occur in the data base. With this choice of a hyperplane, the points x in the feature space that are mapped into the hyperplane are defined by the relation:

$$\Sigma_i \alpha i\, k(xi, x) = \ Constant \tag{12}$$

Note that if K(x,y) becomes small as Y grows farther away from X, each element in the sum measures the degree of closeness of the test point x to the corresponding data base point Xi. In this way, the sum of kernels above can be used to measure the relative nearness of each test point to the data points originating in one or the other of the sets to be discriminated. Note the fact that the set of points X mapped into any hyperplane can be quite convoluted as a result, allowing much more complex discrimination between sets which are not convex at all in the original space.

6 Result Analysis

The classification efficiency of each of the classifiers has been obtained. The results are presented in terms of Sensitivity, accuracy, predicted positive value.

The results for NN, K-NN and SVM are presented in Table. It can be observed from table that for Neural Network, the value of Sensitivity was 100 %, Accuracy was 91.7 % and PPV was found to be 80 %, hence it can be inferred that the accuracy of the system is nearly perfect. The accuracy of the system is comparatively good and can be improved by improving the standard of training along with the number of neurons in the hidden layers (Table 1).

For K Nearest Neighbor, the final value of Sensitivity was 100 %, Accuracy was 83.33 % and PPV was found to be 66 %. The accuracy of the system as compared to Neural Network is less (Tables 2, 3).

Table 1 Final values of the parameters

Parameter	Value	% Value
Sensitivity	1	100
Specificity	0.875	7.5
Accuracy	0.917	91.7
Positive predictive value (PPV)	0.8	80

Table 2 Classification of test images

True positive	True negative	False positive	False negative
4	6	2	0

Table 3 Final values of the parameters

Parameter	Value	% Value
Sensitivity	1	100
Specificity	0.75	75
Accuracy	0.8333	83.33
Predicted positive value (PPV)	0.66	66

Table 4 Classification of test images

True positive	True negative	False positive	False negative
4	7	1	0

Table 5 Final values of the parameters

Parameter	Value	% Valu
Sensitivity	1	100
Specificity	0.875	87.5
Accuracy	0.923	92.3
Predicted positive value (PPV)	0.8	80

Table 6 Comparison of classification

Classifier	Sensitivity (%)	Accuracy (%)	Predicted positive value (%)
Neural Network	100	91.7	80
K NN	100	83.33	66
SVM	100	92.3	80

For Support Vector Machine, the final value of Sensitivity was 100 %, Accuracy was 92.3 % and PPV was found to be 80 %.The accuracy of the system is nearly perfect and slightly better than that of neural network (Tables 4 and 5)

7 Conclusion and Future Scope

The aim of the present paper was analysis of CT images andevaluation of the best classification approach for focal liver lesions of CT image to accurately classify normal and diseased liver.

The preprocessing of CT images was done using anisotropic diffusion filter through which desired output was obtained without much smoothening of the CT image, the image was cropped to remove patient data, followed by seeded region growing technique with manual seed point selection and binary masking to extract liver from CT image was successfully implemented.

Three efficient algorithms to identify and segregate the CT image of liver into diseased and normal has been proposed in this project. This classification was realized by extracting five features from forty-eight images out of which fourteen were diseased images. The features were collected using which codebook was made. This codebook was used to test the accuracy of the dataset using three different methods i.e. Neural Networks, K Nearest Neighbour and Support Vector Machine (Table 6).

Therefore, it can be concluded that Support Vector Machine yields the best results and hence can be used to classify the CT liver images into diseased and Normal.

The algorithms for classification of CT Liver Images can be further improved by the following:

- The algorithm developed in this project is Semi Automated and hence the user has to extract the features and hence this process proves to be time consuming.Saving time using the same algorithm can be done only if the algorithm is made fully automatic.
- The features of the image used for classification are only five, increasing the number of features extracted can improve the efficiency.
- Using a large sample for training and but keeping the ratio of training and testing constant which might improve the performance. Performance is likely to improve with the increase in the size of feature set.

References

1. CC Lee, YC Chiang, CL Tsai, SH Chen (2007) Distinction of liver disease from CT images using kernel-based classifiers. IC-MED, Taiwan
2. Mougiakakou SG, Valavanis IK, Nikita A, Nikita KS (2007) Differential diagnosis of CT focal liver lesions using texture features, feature selection and ensemble driven classifiers. Artif Intell Med 41:25–37
3. Nural W, Yussof JW, Burkhardt H (2009) 3D anisotropic diffusion for liver segmentation. World Acad Sci, Eng Technol 57
4. Susomboon R, Raicu DS, Furst J A hybrid approach for liver segmentation. Intelligent multimedia processing laboratory school of computer science, telecommunications and information systems DePaul university, Chicago, USA
5. Huang Y, Wang L, Li C (2008) Texture analysis of ultrasonic liver image based on wavelet transform and probabilistic neural network. Int Conf BioMed Eng Inform, China
6. Shiraishi J, Sugimoto K, Moriyasu F, Kamiyama N, Doi K (2008) Computer-aided diagnosis for the classification of focal liver lesions by use of contrast-enhanced ultrasonography. Am Assoc Physicists Med 0094-2405(2008)/351734/13

7. Mittal D, Kumar V, Saxena SC, Khandelwal N, Kalra N (2011) Neural network based focal liver lesion diagnosis using ultrasound images. Comput Med Imaging Graph 35:315–323
8. Gletsos M, Mougiakakou SG, Matsopoulos GK, Nikita KS, Nikita AS, Keleki D (2003) A computer-aided diagnostic system to characterize CT focal liver lesions: design and optimization of a neural network classifier. IEEE Trans Inf Technol Biomed 7(3), Sept 2003
9. Singh M, Singh S, Gupta S (2010) A novel algorithim for liver classification using texture analysis. Int Conf Biomed Eng Assistive Technol (BEATS), Dec 2010
10. Harrison D, Rubenfeld DL (1990) Artificial neural networks. SIAM, 23(5):1, 18

Surface Defect Detection of Rubber Oil Seals Based on Texture Analysis

S. Shankar Bharathi, N. Radhakrishnan and L. Priya

Abstract The inspection of surface texture is an important part of many industrial quality control applications. The detection of features and classification based on it in a digital image is the key requirement in quality control systems in production and process. An inspection system to replace human inspectors should be capable of detecting flaws such as scratches, stains, (textural defects) and dents, cracks, blow holes (structural defects) occurring in various shapes and sizes. This paper aims at surface defect detection on rubber oil seals based on texture analysis. The proposed method is based on statistical method by extraction of textural features computed from Gray level co-occurrence matrix with different spatial relationships. As the defects were locally concentrated, computing the textural features of the entire image did not prove to be effective, hence the images were divided and then the features were obtained. Also a unique preprocessing method has been proposed and implemented.

Keywords Surface inspection · Textural abnormalities · Grey-level co-occurrence matrix

S. Shankar Bharathi (✉) · N. Radhakrishnan · L. Priya
TIFAC CORE in Machine Vision, Rajalakshmi Engineering College,
Chennai, Tamilnadu, India
e-mail: shankarbharathi.s@rajalakshmi.edu.in

N. Radhakrishnan
e-mail: radhakrishnan.n@rajalakshmi.edu.in

L. Priya
e-mail: priya.l@rajalakshmi.edu.in

Mohan S. and S. S. Kumar (eds.), *Proceedings of the Fourth International Conference on Signal and Image Processing 2012 (ICSIP 2012)*, Lecture Notes in Electrical Engineering 222,
DOI: 10.1007/978-81-322-1000-9_19,

1 Introduction

Visual inspection for texture has application on a variety of surfaces e.g. wood, steel, rubber, ceramics, and leather, and is highly demanded by industry in order to replace the subjective and repetitive process of manual inspection. Texture analysis based algorithm plays important role in many machine vision application such as surface inspection, scene classification and shape determination. Texture is characterized by spatial distribution of gray levels in a neighborhood. The present techniques used to inspect textural abnormalities are classified into four categories, statistical approaches, structural approaches, filter based methods, and model based approaches [1]. This paper focuses on surface inspection of oil seals based on statistical approaches. Images with different defects such as blemishes, blow holes, cut marks and dents were chosen. The defects could be classified into two categories, textural and structural. Defects like blemishes, small blow holes were considered as textural defects and defects like cut marks, overlap and tear were categorized as structural.

The defects were seen as uneven distribution of gray levels and also as large tonal variation. As the defects were textural abnormalities, determining the textural features and detecting the defects based on it was found to be the appropriate approach. Textural features such as, contrast, homogeneity, energy were computed from Gray Level Co-Occurrence Matrix (GLCM) [3, 5] obtained from the images. As the defects were locally concentrated, the images were divided into smaller images of size 50×50 and GLCM were obtained for each of the subdivided images. GLCM's were obtained with only one spatial relationship, i.e., along 0 degree.

However, single spatial relationship sometimes is not enough to describe the textural features of an image hence, multiple spatial relationships is required. While calculating the statistics we can take the average of GLCMs obtained from it.

Multiple GLCMs were created by specifying four directions, 0 degree, 45 degrees, 90 degrees, and 135 degrees. Features obtained were then averaged. Section 5 shows the difference between both the approaches.

The rest of the paper is structured as follows: Sect. 2 introduces about the machine vision. Section 3 describes a different technique used in preprocessing of the images. In Sect. 4, a brief description of GLCM followed by extraction of textural features is given. In Sect. 5, GLCM with different spatial relationships are described and a comparative study has been shown.

2 Machine Vision

A Machine vision system [4] can be offline or online. In offline, the inspection system is carried out after the production stage i.e. after the component is manufactured. Here the components are let into machine vision system and it is segregated as OK and NOT OK parts.

In other case, machine vision inspection system is installed in the production stage itself. The system checks for the quality of the component just after it is

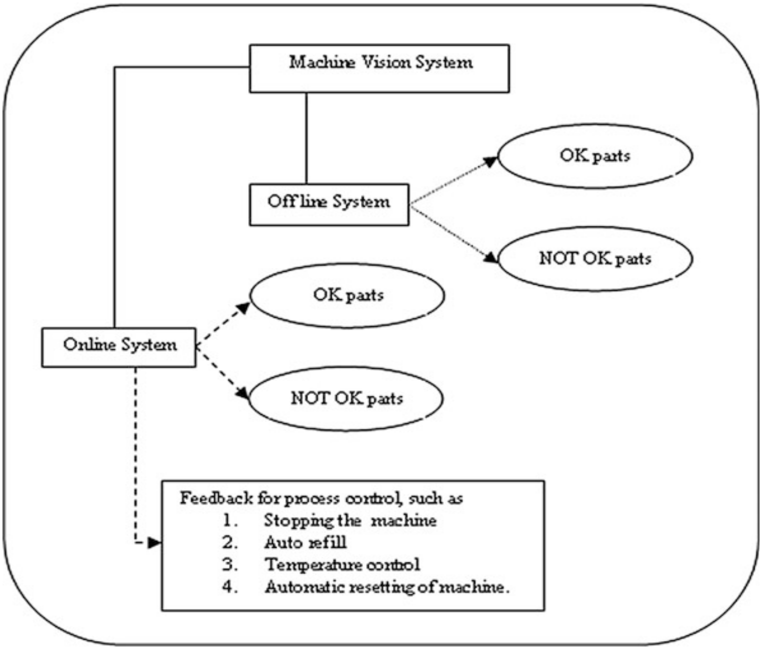


Fig. 1 Online and offline machine vision system

manufactured. In case the system detects many NOT OK components, an alarm signal can be send to the production process immediately for necessary action. This not only reduces the scrap but gives consistent production output. This is the advantage of online inspection system, i.e. giving immediate feedback, which is not the case in offline system. Figure 1 shows the online and offline inspection system.

The oil seals being manufactured were required to be free from any surface defects like blow holes, cut marks, blemishes and overlap. Blow holes, blemishes and overlap occur during the manufacturing process, whereas cut marks occur due to improper handling. The former defects can be controlled while production itself. An online machine vision inspection system can give an alarm or control the manufacturing process before the next lot is manufactured thereby reducing the scrap. A machine vision system should be employed where prevention of failure or the cost of failure is a priority.

3 Preprocessing

Figure 2 shows the images of different defects on the surfaces of the oil seal.

Defects such as blow holes, crack, overlap and blemishes were the noted defects. A texture analysis based algorithm [7], which is proposed here, had to

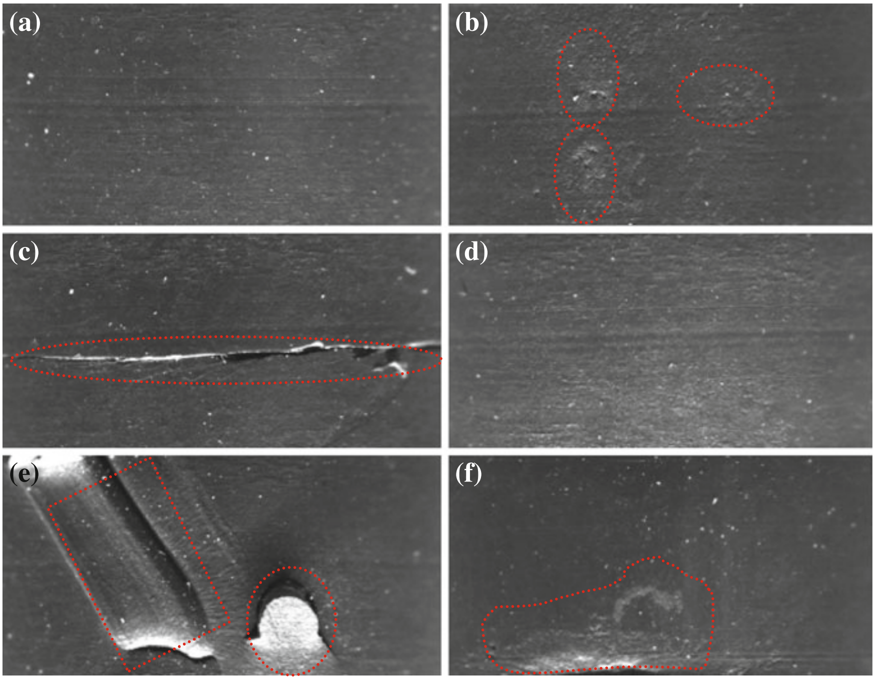


Fig. 2 Example defects, the mark shows the location of defects **a** No defects, **b** blow holes, **c** crack, **d** blemishes, **e** overlap, **f** No defects

detect these defects, but along with these defects the presence of dust particle was also of equal concern, applying the texture analysis algorithm directly to the images without removing the dust particles would give false results, which is undesirable. In general, mean or median filter is used for the removal of dust particles.

But this technique usually blurs the image due to which the information is lost and this may result in false detection while applying the texture analysis algorithm or sometime do not detect the defects. Figure 3 show the application of median

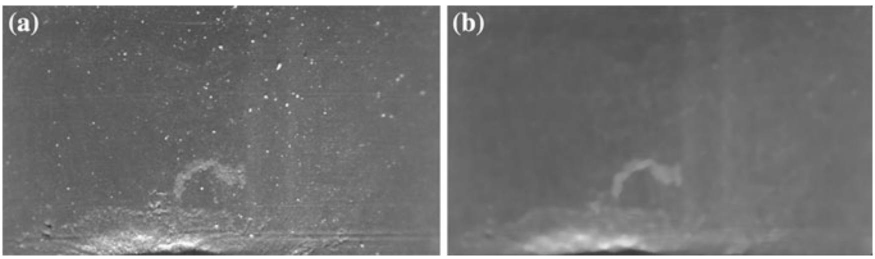


Fig. 3 **a** Original image with blemishes and dust particle. **b** Result of median filter

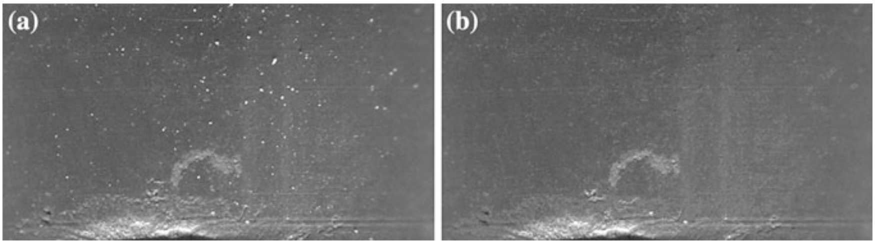


Fig. 4 Result after filtering the dust particle. **a** Original image with blemishes and dust particle. **b** Image after removing the dust particle

filter to one of the images. A unique technique without changing the image other than dust was employed. The technique here used to remove the dust particle was similar to how we perceive the image. In the image it is seen that the dust particle was nothing but foreground. By wiping out the dust particle we would find the background that is nothing but the texture of the rubber. Hence, by changing the gray values of dust particle to surrounding background we are actually removing the dust particle from the image. As each dust particle was filled with different value which dependent on the surrounding area, the imaged looked more uniform. This showed that dust particle was neatly removed without affecting the image. Figure 4 shows the application of this technique.

4 Gray Level Co-occurrence Matrix

The defects found on the surfaces of oil seal were more of textural abnormalities, a bad textured components showed tonal variation and as well as non uniform distribution of it. Defects like blemishes showed increase in tonal unlike in defects like blow holes and cut marks which showed almost zero intensity, also the defects were not uniform but locally concentrated. Global technique such as thresholding in order to segment the defects was not found to be suitable, since defects like blow holes will never be detected. Figure 5 shows how the defects were not segmented out by using threshold technique.

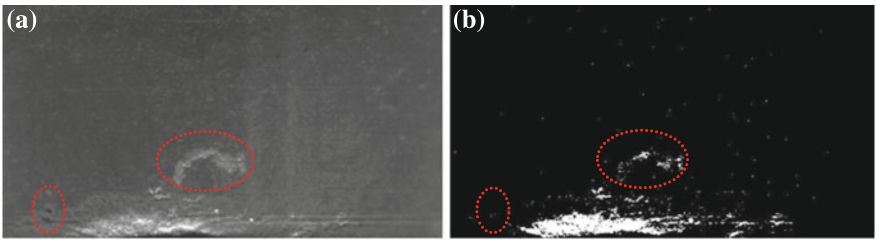


Fig. 5 Result of thresholding **a** Original image with blemishes and blow holes, **b** blemishes were not detected completely and blow holes were not at all detected

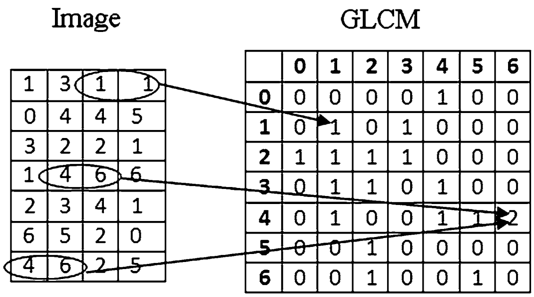


Fig. 6 Formation of GLCM

Hence, texture analysis based method was found to be the appropriate approach. Here statistical based texture analysis method was employed. The statistical approach involved creation of Gray level co-occurrence matrix (GLCM) and obtaining the features out of it. The following gives the brief description of it. GLCM is a statistical method of examining texture that considers the spatial relationship of pixels. The GLCM functions characterize the texture of an image by calculating how often pairs of pixel with specific values and in a specified spatial relationship occur in an image, creating an GLCM, and then extracting the statistical measures from the matrix. To illustrate, fig. 6 shows how GLCM is created.

In the output GLCM, element (1,1) contains the value 1 because there is only one instance in the input image where two horizontally adjacent pixels have the values 1 and 1, respectively. GLCM (4,6) contains the value 2 because there are two instances where two horizontally adjacent pixels have the values 4 and 6. Element (1,2) in the GLCM has the value 0 because there are no instances of two horizontally adjacent pixels with the values 1 and 2. This method continues processing the input image, scanning the image for other pixel pairs (i,j) and records the sums in the corresponding elements of the GLCM. The spatial relationship is defined as the pixel of interest and the pixel to its immediate right (horizontally adjacent), but we can specify other spatial relationships between the two pixels, such as vertical (90 deg.), two diagonals (45 deg. and 135 deg.). The size of the GLCM depends on the bit depth of an image. If an image is of 8 bit depth, (that is intensity range between 0 and 255) the size of the matrix is of 256×256.

The gray-level co-occurrence matrix can reveal certain properties about the spatial distribution of the gray levels in the texture image. After creating the GLCM, there is still one step to take before texture measures can be calculated. The measures require that each GLCM cell contain not a count, but rather a probability. The simplest definition of the probability of a given outcome is the number of times this outcome occurs, divided by the total number of possible outcomes. This process is called normalizing the matrix. After creating the normalized GLCM, several statistics can be derived. These statistics provide information about the texture of the image. Here we have only used contrast to detect the defects. The reason for not choosing other features for detecting the defects is explained later in this section. The following explains how contrast is calculated.

Table 1 Textural features derived for all the images

Property	Image 2a	Image 2b	Image 2c	Image 2d	Image 2e
Contrast	32.2781	28.4855	41.1798	47.4830	89.7106
Correlation	0.8103	0.8913	0.7700	0.9315	0.9761
Energy	0.0028	0.0033	0.0036	0.0023	0.0016
Homogeneity	0.3995	0.4416	0.3926	0.3785	0.3826

Values on the GLCM diagonal show no contrast, and contrast increases away from the diagonal. So we have to create a weight that increases as distance from the diagonal increases. The following gives the contrast equation.

$$\sum_{i,j=0}^{N-1} P_{ij}(i-j)^2 \tag{1}$$

where $P_{i,\,j}$ is normalized GLCM value at i,j th location. When i and j are equal, the cell is on the diagonal and (i-j) = 0. These values represent pixels entirely similar to their neighbour, so they are given a weight of 0. If i and j differ by 1, there is a small contrast, and the weight is 1. If i and j differ by 2, contrast is increasing and the weight is 4. The weights continue to increase exponentially as (i-j) increases. Table 1 shows the statistics derived from GLCM created for each of the image shown above based on east relationship, i.e., the neighbor pixel was chosen right side to reference pixel. Other spatial relationships were also considered. A comparison was made by choosing all the spatial relationships and averaging it. The figures in Sect. 5 show it. The contrast value for image 2d and e are considerably higher compare to 2a.

The difference may be sufficed to differentiate between a good textured rubber and a bad one. But for image 2b and c the values are nearer to 2a. Other features like correlation, energy, homogenity also did not show much of differences. As the defects were locally concentrated finding the GLCM for the entire image did not prove to be effective and accurate. In order to detect the defects much accurately the image was divided into smaller images and properties were extracted from it. The images were resized such that entire image was divided equally and each subdivided images were of equal size. The image was divided into 72 sections, with each section of size 50 × 50. For each section the GLCM with only one spatial relationship was found and the properties were determined. From all the features extracted only contrast showed very high differences between the images. Table 2 shows the maximum and minimum of all the features extracted from GLCM obtained for 72 sections from each of three images. As seen the maximum correlation for images 2a, b and c were almost same.

The other features did show some differences but were very close unlike values of contrast which are very high; hence contrast was chosen to identify the defects.

The following images show the detected defects (Fig. 7).

Table 2 Maximum and minimum of properties

Properties	Image 2a	Image 2b	Image 2c
Maximum contrast	101.4759	182.7053	186.2665
Minimum contrast	4.9510	2.6147	1.5555
Maximum correlation	0.8926	0.9036	0.9503
Minimum correlation	0.3839	0.4743	0.6208
Maximum energy	0.0193	0.0283	0.0323
Minimum energy	0.0025	0.0015	0.0012
Maximum homogeneity	0.5709	0.6437	0.6908
Minimum homogeneity	0.2968	0.2283	0.2171

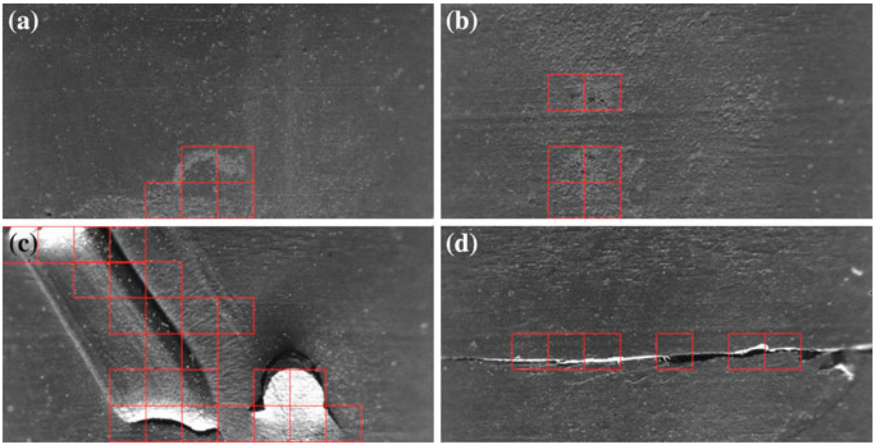


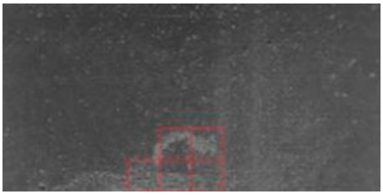
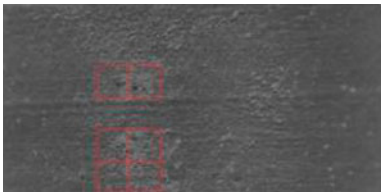
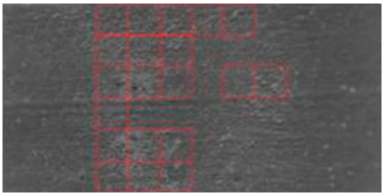
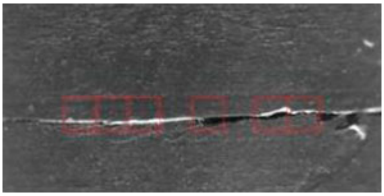
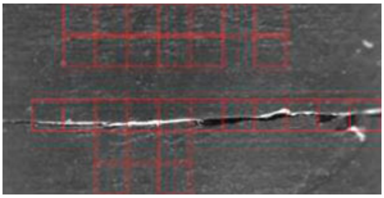
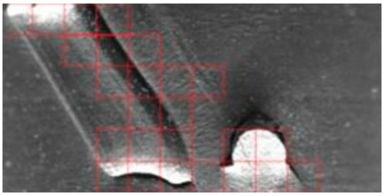
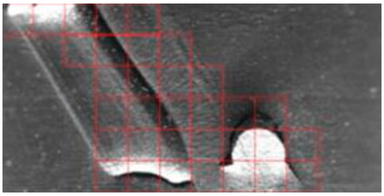
Fig. 7 Rectangular boxes shows the defects detected using single spatial relationship

It shows that the above method of detecting defect using GLCM with single spatial relationship is effective; however it did not prove to be accurate. As it is seen in the above images, some of the defects were not detected.

5 GLCM with Different Spatial Relationships

Single spatial relationship sometimes is not enough to describe the textural features of an image hence, multiple spatial relationships is required. While calculating the statistics we can take the average of GLCMs obtained from it. Multiple GLCMs were created by specifying four directions, 0 degree, 45 degrees, 90 degrees, and 135 degrees [8]. Features obtained were then averaged. Table 3

Table 3 Comparison between single and multiple spatial relationships

Defects detected using single spatial relationship	Defects detected using Multi spatial relationship
	
	
	
	

shows the difference in detection of defects by choosing single and multiple spatial relationships. It showed that by employing multiple spatial relationships the defect detection was improved, coarse texture was also detected.

6 Results

The algorithm was developed using Image Processing Toolbox in MATLAB. Around 100 images having all defects described in this paper were taken for study. The results were found to be satisfactory. Application of texture analysis with single spatial relationship locally rather than applying fully on the image, i.e. on the subdivided images, improved the defect detection but, only 70 % of the defects in the image were detected. However, it improved to 96 % by choosing multiple spatial relationships.

7 Conclusion and Future Work

This paper has been focused on developing a methodology and implementation of texture analysis based algorithm to detect defects on rubber oil seals. The work involved: (1) Generation of Gray level co-occurrence matrix; (2) Extraction of textural features from the GLCM. Textural features were obtained from GLCM with single and as well as with multiple spatial relationships. In order to detect the defects accurately the images were divided and textural features were derived. As a future enhancement, a new algorithm will be developed by improving the existing one by involving different technique to create GLCM and extraction of features. However, from the work undertaken to date, it is clear that texture based analysis alone is not sufficient in order to detect the defects.

References

1. Xianghua Xie.: A Review of Recent Advances in Surface Defect Detection using Texture analysis Techniques. Electronic Letters on Computer Vision and Image Analysis 7(3):1-22, 2008
2. Jonathan M Blackledge and Dmitry A Dubovitskiy.: A Surface Inspection Machine Vision System that Includes Fractal Texture Analysis. ISAST Transactions on Electronics and Signal Processing
3. R. Haralick, K. Shanmugan, and I. Dinstein.: Textural features for image classification. IEEE Transactions on Systems, Man, and Cybernetics. 3(6):610–621, 1973
4. TANG Bo, KONG Jian-yi, WANG Xing-dong and Chen Li.: Surface Inspection System of Steel Strip Based on Machine Vision. First International Workshop on Database Technology and Applications, 359–362, 2009
5. Sina Jahanbin, Alan C. Bovik, Eduardo P'erez,DineshNair.: Automatic Inspection of Textured Surfaces by Support Vector Machines
6. T. Caelli and D. Reye.: On the classification of image regions by colour, texture and shape. Pattern Recognition, 26(4):461–470, 1993
7. J. Chen and A. Jain.: A structural approach to identify defects in textured images. In IEEE International Conference on Systems, Man, and Cybernetics, volume 1, pages 29–32, 1988
8. http://www.fp.ucalgary.ca/mhallbey/tutorial.htm

Channel Estimation and Equalization for Time-Varying OFDM System Using Kalman Filter

C. Rajasekhar, D. Srinivasa rao and K. M. K. Chaitanya

Abstract Orthogonal Frequency Division Multiplexing (OFDM) technique has become major multi carrier technique with the advent of wireless and mobile communications. It is known that OFDM is effective over frequency selective fading. However, ICI (inter-carrier interference) destroys the subcarrier orthogonality in time varying channels. The ICI can be mitigated by using different channel estimation schemes at receiver of fast fading systems.Though many channel estimation techniques have already existed, they suffer from complexity in implementation. Efficient channel estimation can be done by assuming channel as a linear state space model i.e., Kalman filter model. This paper discusses about the Kalman filter based channel estimation technique that can be implemented with less complexity yet efficient for a time varying channel model. The Equalization is employedusing a time-domain filter that maximizes receiver SINR. Also.the proposed technique's efficiency is proved in terms of MSEE, BER in comparison with other techniques.

Keywords OFDM · ICI · Time varying channel · Kalman filter model · Tapped delay line model · CE-BEM channel · Equalization · MSEE · BER

C. Rajasekhar (✉) · K. M. K. Chaitanya
Department of ECE, GITAM University, Vishakapatanam, Andhra Pradesh, India
e-mail: chukkarajasekhar@gmail.com

D. Srinivasa rao
Department of ECE, GMRIT, Rajam, Andhra Pradesh, India
e-mail: srinivasa.dasari@gmail.com

Mohan S. and S. S. Kumar (eds.), *Proceedings of the Fourth International Conference on Signal and Image Processing 2012 (ICSIP 2012)*, Lecture Notes in Electrical Engineering 222,
DOI: 10.1007/978-81-322-1000-9_20,

1 Introduction

Orthogonal frequency division multiplexing is extensively used in wireless communication systems which involve high data rates. However, these systems are highly prone to time selective fading which affects subcarrier orthogonality and introduces ICI. Review of recent literature emphasizes that use of OFDM over rapid time varying systems is limited by ICI and channel state information. The channel estimation techniques available in literature are not capable of providing reliable estimates-with low overhead and channel is rapidly varying with time. This paper discusses channel estimation technique in high mobility wireless systems that considers channel to be linear state space model a Kalman Filter. The channel is modeled using BEM and Kalman based estimation is carried out. Equalization of the estimated channel is implemented by maximizing the SINR. The proposed scheme is compared with conventional training method and show that the technique outperforms in bandwidth as well in BER performance.

2 OFDM System Model

The block diagram of the proposed system is shown in Fig. 1. The OFDM system is based on Cyclic Prefix. In the transmitter, the data is modulated with QPSK that uses 2 bits for a symbol. The modulated frequency domain signal is converted into time domain by the use of Inverse Fast Fourier Transform (IFFT). IFFT allocates the single data stream to multiple subcarriers, and it also assures that the subcarriers are orthogonal to each other.

The disturbed data is given to Serial to parallel converter by discarding the cyclic prefix. The parallel data is Channel Estimated with the proposed model and is converted to frequency domain by the use of FFT. The transformed data is demodulated by QPSK.The obtained data is free from Inter Symbol Interference

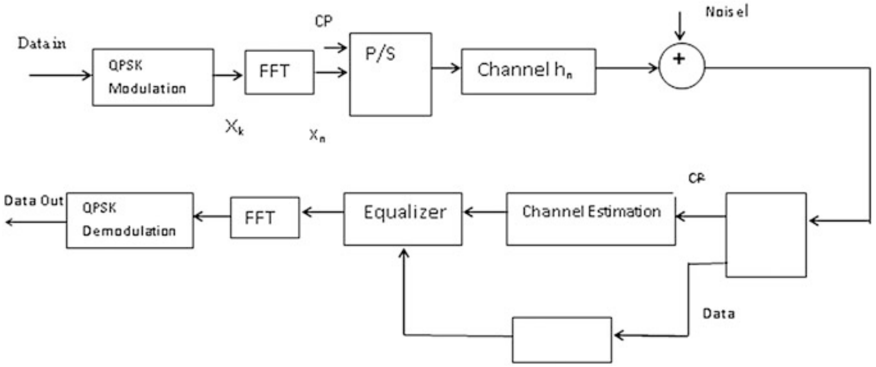


Fig. 1 OFDM system model

with the use of orthogonal subcarriers and free from Inter Carrier Interference (ICI) with the use of cyclic prefix.

2.1 Channel Model

The fading channel is assumed to be selective in both time and frequency domains with wide sense stationary uncorrelated scattering (WSSUS). To cater the frequency selectivity, the channel impulse response is implemented as tapped delay line model and CE–BEM channels for comparison, where the individual taps are i.i.d (independent and identically distributed).

The transmitted signal can be given by the equation:

$$y[n] = hx[n] + z[n]. \tag{1}$$

For complex exponential basis exponential model (CE-BEM) (Fig. 2)

$$h(n, l) = \sum_{q=-Q/2}^{Q/2} c(l, q) e^{jw_q n}. \tag{2}$$

where c (l, q) are the complex exponential basis which can be defined by the matrix:

$$\mathrm{C(1, q)} = \begin{bmatrix} 1 & \dots & 1 \\ \exp\left(-j\omega \frac{Q}{2}\right) & \dots & \exp\left(j\omega \frac{Q}{2}\right) \\ \vdots & & \vdots \\ \exp\left(-j\omega n \frac{Q}{2}\right) & \dots & \exp\left(j\omega n \frac{Q}{2}\right) \end{bmatrix}.$$

And,

$$w_q = \frac{2\Pi q}{G(N + gi)}, q = -\frac{Q}{2}, \ldots\ldots 0, 1, \ldots\ldots \frac{Q}{2}. \tag{3}$$

G is the factor denoting the number of samples in Doppler frequency domain, and the number of coefficients per tap is Q +1 where Q is determined as

$$Q = 2[f_D T_s (N + \mathrm{g}i)\, G]. \tag{4}$$

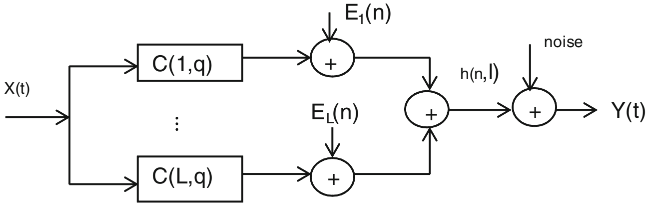


Fig. 2 Complex exponential—basis expansion (CE-BEM) channel model

The relation, between the number of coefficients and the sampling factor, is obtained by considering that for a close match between the Jakes' spectrum and the BEM, the highest frequency of the BEM, which is observed from (3) to be Q/(2G(N + gi)Ts) should be close to the highest Doppler frequency f_D The received symbol under the effect of time varying fading channel and AWGN is given by

$$y(n) = \sum_{l==0}^{L-1} h(n,l)x_{n-1} + z(n). \tag{5}$$

The above equation may be written in vector form as:

$$\mathrm{Y} = \bar{H}\mathrm{Q}^H\mathrm{X} + \mathrm{Z}. \tag{6}$$

where Q is the standard N-point DFT matrix to diagnolize the matrix, y, z are N × 1 vectors of received symbol and AWGN samples respectively, $\bar{\mathrm{H}}$ is an N × N matrix comprising the channel taps at each instant of time over the entire OFDM symbol.

$$\bar{H} = \begin{bmatrix} \mathrm{h}(0,0) & 0 & \ldots & \mathrm{h}(0,L-1) & \ldots & \mathrm{h}(0,1) \\ \mathrm{h}(1,1) & \mathrm{h}(1,0) & \ldots & \ldots & \ldots & \mathrm{h}(1,2) \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ \vdots & \vdots & \vdots & \vdots & \mathrm{h}(N-2,0) & 0 \\ 0 & 0 & & & \mathrm{h}(N-1,1) & \mathrm{h}(N-1,0) \end{bmatrix}$$

This matrix representation takes into account the effects of CP as well. It considers the OFDM system to be time invariant over a block of data.

3 Channel Estimation

Channel estimation is a technique, in which the channel state information is extracted using the channel impulse response. By considering channel as a linear order Kalman filter model, provides us with the basis for defining a state-space model, with (7) and (8) as the process equation and measurement equation respectively.

$$Y(n) = \bar{X}_n S_n + z(n). \tag{7}$$

$$\mathrm{S_n} = \mathrm{AS_{n-1}} + \mathrm{V_n}. \tag{8}$$

Here, the unknown state at instant n, Sncomprises of 'L' TDL coefficients within the multi frame.

$$\mathrm{S_n} = \begin{bmatrix} \mathrm{c}(0,Q) \\ \mathrm{c}(1,Q) \\ \\ \mathrm{c}(\mathrm{L}-1,Q) \end{bmatrix}_{\mathrm{LX1}} \tag{9}$$

$$\bar{X}_n = \begin{bmatrix} X_n e^{-jw_Q n} \\ X_{n-1} e^{-jw_Q n} \\ \ddots \\ X_{n-L+1} e^{-jw_Q n} \end{bmatrix}^T. \tag{10}$$

Measurement matrix is a $1 \times L$ vector, which uses the L transmitted symbols that affect the received symbol at instant n. In case of CE-BEM channel, the unknown state at instant n, Sncomprises of L(Q +1) BEM coefficients within the multi frame.

$$S_n = \begin{bmatrix} c\left(0, -\frac{Q}{2}\right) \\ \vdots \\ c(0,0) \\ \vdots \\ c\left(0, \frac{Q}{2}\right) \\ c\left(1, -\frac{Q}{2}\right) \\ \vdots \\ c\left(L-1, -\frac{Q}{2}\right) \\ \vdots \\ c\left(L-1, \frac{Q}{2}\right) \end{bmatrix}_{(Q+1)L+1} \tag{11}$$

Measurement matrix:

$$\bar{X}_n = \begin{bmatrix} X_n e^{-jw_{-Q/2} n} \\ \vdots \\ X_n e^{-jw_0 n} \\ \vdots \\ X_n e^{-jw_{Q/2} n} \\ X_{n-1} e^{-jw_{-Q/2} n} \\ \vdots \\ X_{n-L+1} e^{-jw_{-Q/2} n} \\ \vdots \\ X_{n-L+1} e^{-jw_{Q/2} n} \end{bmatrix}^T. \tag{12}$$

is a $1 \times (Q + 1)L$ vector, which uses the L transmitted symbols that affect the received symbol at instant n.These may be the CP symbols of the current and the previous symbol.The measurement noise vector z(n) comprises the AWGN sample at instant n. Since, the channel coefficients remain invariant over a multi frame; the transition matrix in (8) is simply an identity matrix.

For TDL channel it is,

$$A = I_L \tag{13}$$

Whereas in CE-BEM channel it is,

$$A = I_{L(Q+1)} \tag{14}$$

Whereas the process noise vector comprises all zeros. For TDL, it is

$$V_n = o_{LX1} \tag{15}$$

For CE-BEM it is,

$$V_n = o_{L(Q+1)X1} \tag{16}$$

4 Simulation Results

The following are the parameters being considered for the simulation of Tapped Delay Line and CE-BEM channel Models and Kalman filter model based channel estimation and Equalization.

Parameter	Value
FFT size	128
Subcarriers ‘N’	128
Channel Taps ‘L’	4
Cyclic prefix ‘g’	8

The system is evaluated at different Doppler spreads, $f_D = 0.01, 0.07, 0.1$ and 0.2. Here we have given results for CP based estimation. Only few of the results are presented (Figs. 3, 4, 5, 6).

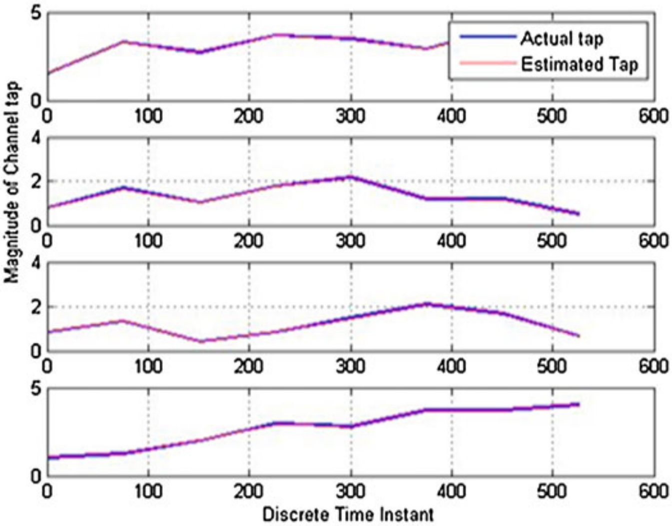


Fig. 3 CP based kalman filter channel estimation for CE-BEM channel

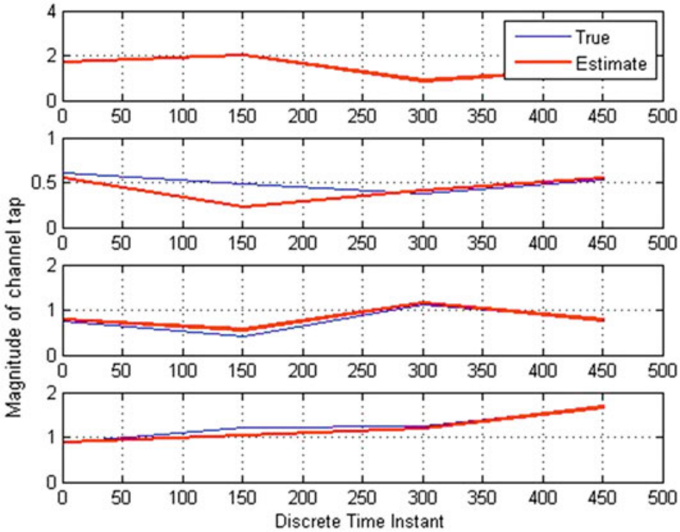


Fig. 4 CT based Kalman filter channel estimation for CE-BEM channel

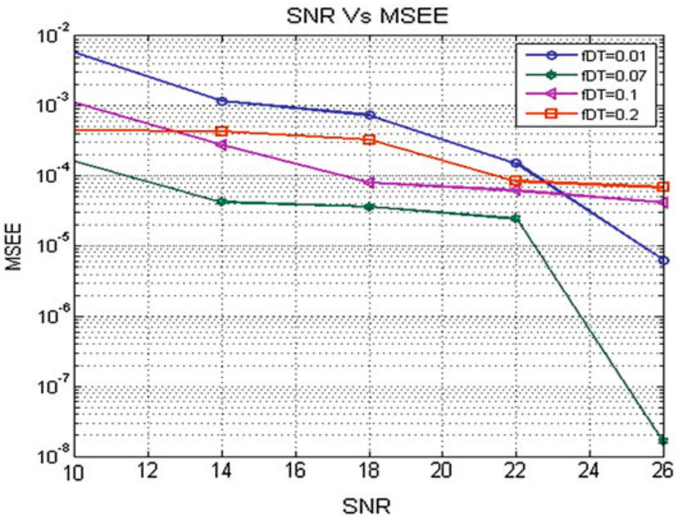


Fig. 5 SNR Vs MSEE at different doppler spreads for CP based estimator

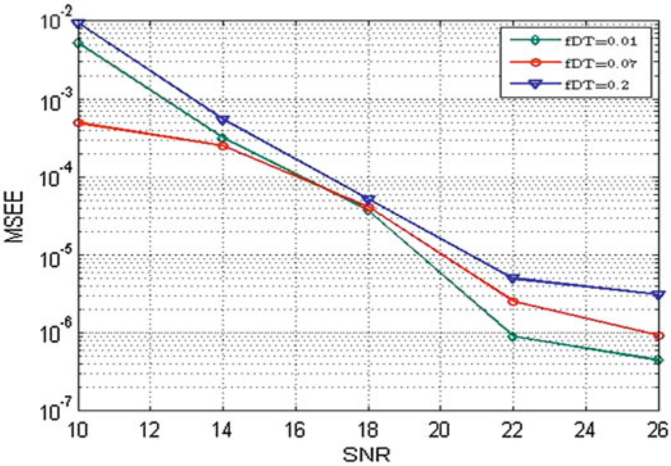


Fig. 6 SNRVs MSEE at different doppler spreads for CT based estimator

5 Conclusion

The Time Varying channel model is designed as TDL and CE-BEM. The channel estimation is done by considering channel as Kalman filter Model. The proposed channel estimation scheme is compared with conventional scheme involves training sequences. The proposed Equalization scheme is evaluated in terms of

SINR at different Doppler spreads. The whole OFDM system performance is evaluated in terms of BER and PAPR is evaluated. The simulation is carried out using MATLAB Software and from the simulation results we can conclude that the CE-BEM channel model well suits for the proposed estimation scheme and performs well in high mobility OFDM system in terms of MSEE, SINR and BER.

References

1. van de Beek JJ, Edfors O, Sandell M,Wilson SK, Borjesson PO (1995). On channel estimationin OFDM systems. In: Proceedings IEEE 45th Vehicle Technology Conferences Chicago, pp 815–819
2. Negi R, Cioffi J (1998) Pilot tone selection for channel estimation in a mobile OFDM system. IEEETrans Consumer Electron 44(3):1122–1128
3. Yuping Z, Aiping H (1997) A novel channel estimation method for OFDM mobile communicationsystems based on pilot signals and transform-domain processing. In: The 47th IEEE Vehicle Technology Conferences, vol 4–7. pp 2089–2093
4. Gupta P, Mehra DK (2006). Kalman filter based channel estimation and ICI suppression for highmobility OFDM systems. In: The IEEE international conferences on industrial Technology, ICIT 2006 15–17, pp 612–615
5. Mazet L, Buzenac-Settineri V, de Courville M, Duhamel P (2002) An EMbased semi-blind channel estimation algorithm designed for OFDM systems. In: the 36th asilomar conferences on signals. Syst Comp 2:1642–1646
6. Ma X, Kobayashi H, Schwartz SC (2002). An EM-based estimation of OFDM signals. In: IEEE wireless communication and networking Confercences, WCNC-2002, vol 1. pp 228—232). 17–21 Mar 2002
7. Moon TK (1996) The expectation-maximization algorithm. IEEE Sig Proc Mag 13(6):47–60
8. Muquet B, de Courville M, Duhamel P (2002) Subspace-based blind and semi-blind channel estimation for OFDM systems. IEEE Trans Signal Process 50(7):1699–1712
9. Petropulu A, Ruifeng Z, Lin R (2004) Blind OFDM channel estimation through simple linear precoding. IEEE Trans Wireless Commun 3(2):647–655
10. Wang X, Liu KJR (1999) Adaptive channel estimation using cyclic prefix in multicarrier modulationsystem. IEEE Commun Lett 3(10):291–293
11. Wang X, Liu KJR (2003) Performance analysis for adaptive channel estimation exploiting cyclic prefix in multicarrier modulation systems. IEEE Trans Commun 51(1):94–105
12. Tsatsanis MK, Giannakis GB (1996) Modeling and equalization of rapidly fading channels. International Journal of Adaptive Control Signal Processing. Wiley 10:159–176
13. Giannakis GB, Tepedelenlioglu C (1998) Basis expansion models and diversity techniques for blind identification and equalization of time-varying channels. Proc IEEE 86(10): 1969–1986
14. Ma X, Giannakis GB, Ohno S (2003) Optimal training for block transmissions over doubly selective wireless fading channels. IEEE Trans Signal Process 51(5):1351–1366
15. Haykin S (2002). Adaptive filter theory, 4th edn. Pearson Education, London, pp 466–485

A Remote Healthcare Monitoring System for Faster Identification of Cardiac Abnormalities from Compressed ECG Using Advanced Data Mining Approach

N. Sathiya Rani, K. Vimala and V. Kalaivani

Abstract Cardiac Disease has become very common perhaps because of increasingly busy lifestyles. The rapid advancement of mobile communication technologies offers innumerable opportunities for the development of software and hardware applications for remote monitoring of chronic disease. This paper describes a remote health-monitoring service that provides an end-to-end solution. We present an efficient data mining-based solution that recognizes different CVDs (such as ventricular flutter/fibrillation, atrial fibrillation, atrial premature beat, premature ventricular contraction) from the compressed ECG, it was proposed to perform real-time classification of Cardiac Vascular Disease (CVD) based on data mining techniques. The subset of the features selection from the compressed ECG was performed using the Genetic algorithm and the clustering was performed using Expectation Maximization.

Keywords Compressed ECG · Genetic algorithm · Clustering · Rule based Technique

N. Sathiya Rani (✉) · K. Vimala · V. Kalaivani
Department Computer Science and Engineering, National Engineering College, Kovilpatti, India
e-mail: n.sathiya90@gmail.com

K. Vimala
e-mail: vimala.arivalagan@gmail.com

V. Kalaivani
e-mail: kalai.engr@rediffmail.com

Mohan S. and S. S. Kumar (eds.), *Proceedings of the Fourth International Conference on Signal and Image Processing 2012 (ICSIP 2012)*, Lecture Notes in Electrical Engineering 222, DOI: 10.1007/978-81-322-1000-9_21, © Springer India 2013

1 Introduction

Electrocardiogram (ECG) provides the detailed condition of the heart of a cardiac patient. ECG also contains some features, which can serve as a biometric entity for identification of a particular patient. It is well known that modern clinical systems require the storage, processing and transmission of large quantities of ECG signals. The aim of ECG compression algorithms is to achieve a reduced information rate, while retaining the relevant diagnostic information in the reconstructed signal. Efficient and low computationally complexity compression schemes for medical signals are useful in applications related to mobile healthcare and real-time patient monitoring but also in optimized databases.

Advanced monitoring solutions using telecommunicating technologies are used for remote ECG diagnosis. The use of telecommunications for remote diagnosis is growing rapidly, and there are several products and projects within mobile ECG recording using Internet solutions, Bluetooth technology, cellular phones, WAP-based implementations and wireless local area networks, WLAN. A remote diagnosis system integrating digital telemetry has been developed, using a wireless patient module, a homecare station and a remote clinical station. Traditionally 24/72 h ECG-recording systems like “Holter-monitoring” can today use built-in mobile telephones to send information to the hospital, but is mostly used with a recording unit that physically has to be carried to the hospital for analyzes.

2 Related Works

There are many researches are going in analyzing ECG signals in national and international level. For example, the telemedicine facilities have been developed for diagnosis through the Internet. Usually attribute selection [1] is done for selecting the features from the compressed ECG. Expected Maximization (EM) [2] clustering technique is used to create normal and abnormal ECG clusters. The threshold value of ECG signal determination is done using Wavelet transform coefficients [3].

For Remote Health Care monitoring, the compression of ECG signals are essential. For compression new wavelet-based [4] ECG compression technique is implemented and this technique is tested using several records taken from the MIT-BIH Arrhythmia database. Also, wavelet transform method is used to achieve high compression ratios (CR) and a low percent root mean square difference (PRD). Using smartphone-based wearable CVD-detection platform, real-time ECG acquisition and display, feature extraction, and beat classification are done.

Noise is present usually in ECG pattern and the noise reduction process is an important task in ECG signals. Such noises are difficult to remove using typical filtering procedures. So many noise reduction technique is developed in international level such as efficient analytical tool [5] which is a technique for averaging of cardiac cycles which is increasing signal to noise ratio.

Usually the parameters are given as the inputs from medical practitioner directly on to the mobile phone. The score [6] is calculated by miniature java based software running inside the mobile phone. Based on the score, level of urgency is determined by the intelligent program. At the end, specialists are contacted automatically by messaging services. Moreover, the results of the scoring are transmitted to the hospital server.

The small capacitive electrodes integrated into a cotton T-shirt together with a signal processing and transmitting board on a two-layer standard printed circuit board design technology. The entire system [7] has small size, is thin, and has low power consumption compared to recent ECG monitoring systems. In addition, appropriate signal conditioning and processing is used to remove motion artifacts. The acquired ECG signals are compared with the obtained using conventional glued-on electrodes, and are easily read and interpreted by a cardiologist.

Prototype Intelligent Heart Disease Prediction System (IHDPS) [1] are developed in national level using data mining technique such as Decicion trees, Navie Bayes, Neural Network. Also,biometric systems have been developed using ECE using data mining (DM) techniques like attribute selection and clustering. The biometric template has lesser in size compared to other forms of biometrics like face, finger, retina, etc.

The ECG signal is filtered using digital filtering techniques to remove power line interference and base line wander. Support Vector Machine(SVM) [8] is used as a classifier for detection of QRS complexes in ECG.

An Efficient DDC algorithm has been developed over existing modified Amplitude Zone Time Epoch Coding (AZTEC) technique, named as improved modified AZTEC [9] and tested on Common Standard for quantitative Electrocardiography (CSE) database. The performance has been evaluated on the basis of compression ratio (CR), percentroot- mean-square difference (PRD) and fidelity of the reconstructed signal. A comparison of the wavelet- derived features of compressed and original signals has been used for performance evaluation of the compressed signal.

The ultimate aim of this proposed work is to provide a remote healthcare monitoring system for Faster Identification of Cardiac Abnormalities from Compressed ECG using advanced data mining approach. It is proposed to implement the data mining techniques on a hospital server that will generate a set of constraints for representing each of the abnormalities. Then, the patient's mobile phone receives these set of constraints and employs a rule-based system, that can identify each of abnormal beats in real time.

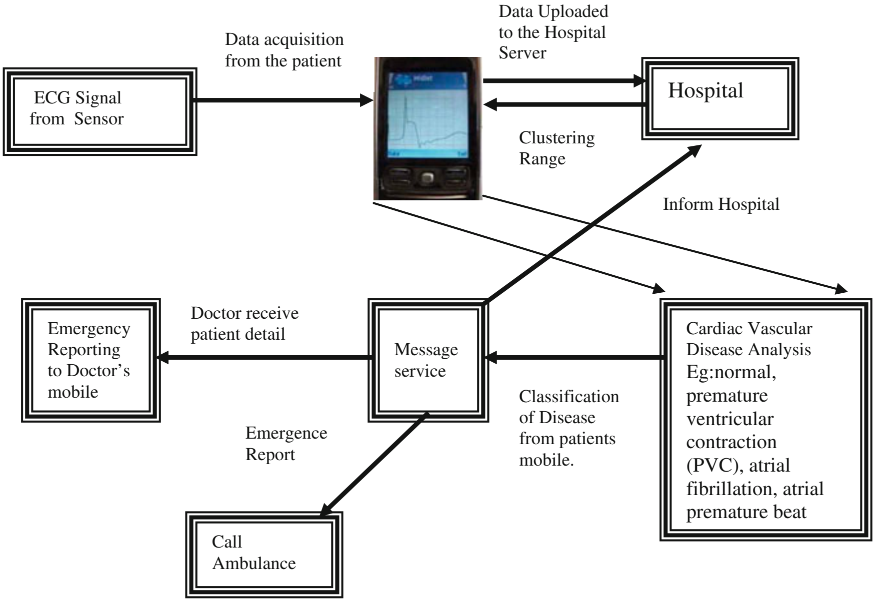


Fig. 1 Cardiavascular abnormality diagnosis process

3 Methodology

This system provides an efficient telecardiology diagnosis system for both patients and doctors for the fast identification and treatment of the disease. Figure 1 shows the representation of cardiavascular abnormality diagnosis process.

3.1 Feature Subset Selection

In the proposed system, Correlation-based Attribute Selection using Genetic Algorithm is used. The feature subset selection process is used to reduce the dimensionality of the analyzed data, it will speed up learning algorithms, improves the performance of data mining techniques (e.g., learning time, predictive accuracy, etc.), and improves the comprehensibility of the output.

The genetic algorithm (GA) is an optimization and search technique based on the principles of genetics and natural selection. A GA allows a population composed of many individuals to evolve under specified selection rules to a state that maximizes the fitness. A genetic algorithm mainly composed of three operators: selection, crossover, and mutation.

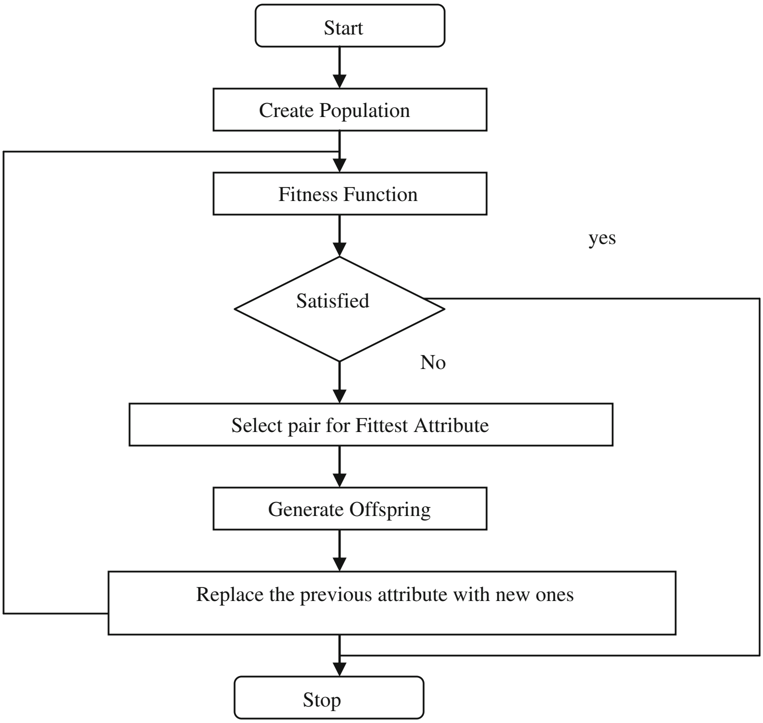


Fig. 2 Process involved in attribute selection

Table 1 Correlation matrix

Chromosome labels	Chromosome frequency	Chromosome strings
A	25	0011001
B	45	0101101
C	75	1001011
D	105	1101001
E	63	0111111
F	85	1010101
G	93	1011101

Proposed GA based algorithm for selection of optimal subset of attributes is diagrammatically represented in figure(Proposed method for selecting optimal subset of attributes). Proposed method can perform either equally good or better than many of the existing methods and accuracy is more when applied on real and large dataset. GA is used to search the optimal subset of attributes besides being used for searching the optimal techniques for attribute selections amongst the available ones (Fig. 2).

Table 2 Chromosome with their fitness value

Chromosome labels	Chromosome frequency	Chromosome strings
A	25	0011001
D	105	1101001
G	93	1011101

To apply the GA attributes are required to be encoded as chromosomes so that the GA operations can be applied on them. Encoding is a process of representing individual genes. The process can be performed using bits, numbers, trees, arrays, lists or any other objects. Bit representation of attributes is done over here (Table 1).

In selection, a good string (on the basis of fitness) is selected to breed a new generation. Fitness function for GA is a simple function, which assigns a rank to individual attribute on the basis of correlation coefficients is evaluated which shall fit to take part in the crossover operations that are having lower correlation coefficients. The fitness function is taken over here is;

$$f(x) = 1 - \min(rx) \tag{1}$$

where, min (rx) is minimum value of correlation coefficient corresponding to any attribute X (Table 2).

We can now produce clusters from the normal as well as different abnormal compressed ECG patterns using the smaller subset of attributes. After the cluster formation task, the hospital server can determine the cluster mean and the cluster ranges. The proposed system works on the compressed ECG character frequency, and does not even require decompression, which would take valuable extra time from the patient's life.

3.1.1 Clustering

The Expectation maximization clustering is performed on the attribute selected data which clusters the data based on disease characteristics, efficient data mining-based solution that recognizes different CVDs (such as ventricular flutter/fibrillation, atrial fibrillation, atrial premature beat, premature ventricular contraction) was estimated.

3.1.2 Initialization

Each class j, of M classes (or clusters), is constituted by a parameter vector (θ), composed by the mean (μj) and by the covariance matrix (P_j), which represents the features of the Gaussian probability distribution (Normal) used to characterize the observed and unobserved entities of the data set x.

$$\theta(t) = \mu_j(t),\ P_j(t),\ j = 1\ldots M \tag{2}$$

On the initial instant (t = 0) the implementation can generate randomly the initial values of mean (μj) and of covariance matrix (P_j). The EM algorithm aims to approximate the parameter vector (θ) of the real distribution. Another alternative offered by MCLUST is to initialize EM with the clusters obtained by a hierarchical clustering technique.

3.1.3 E-Step

This step is responsible to estimate the probability of each element belong to each cluster ($P(C_j|x_k)$). Each element is composed by an attribute vector (x_k). The relevance degree of the points of each cluster is given by the likelihood of each element attribute in comparison with the attributes of the other elements of cluster Cj.

$$P(Cj|x) = \frac{(|\sum_j (t)|^{-1/2} \exp^{nj} P_j(t))}{(\sum_{k=1}^{M} |\sum_j (t)|^{-1/2} \exp^{nj_j} P_j(t))} \tag{3}$$

3.1.4 M-Step

This step is responsible to estimate the parameters of the probability distribution of each class for the next step. First is computed the mean (μ_j) of classe j obtained through the mean of all points in function of the relevance degree of each point.

$$\mu_j(t+1) = \frac{\sum_{k=1}^{N} P(C_j/x_k)x_k}{\sum_{k=1}^{N} P(C_j/x_k)} \tag{4}$$

To compute the covariance matrix for the next iteration is applied the Bayes Theorem, which implies that P(A|B) = P(B|A) * P(A)P(B), based on the conditional probabilities of the class occurrence.

$$\sum_j (t+1) = \frac{\sum_{k=1}^{N} P(Cj|x_k)(x_k - \mu_j(t))(x_k - \mu_j(t))}{\sum_{k=1}^{N} P(C_j|x_k)} \tag{5}$$

The probability of occurrence of each class is computed through the mean of probabilities (C_j) in function of the relevance degree of each point from the class.

$$\mathrm{P_j(t+1)} = \frac{1}{N}\sum_{k=1}^{N} P(C_j|x_k) \tag{6}$$

The attributes represents the parameter vector θ that characterize the probability distribution of each class that will be used in the next algorithm iteration.

3.1.5 Rule Based Technique

Using the proposed technique, one or more diseases can be successfully identified. Γ represents the constraint set for N number of diseases. Each element within Γ (i.e., an individual disease) contains the actual frequency constraints (i.e., character frequency). If a patient is recommended for ongoing monitoring, then during the holter monitoring the doctors could have already obtained the patient's normal and multiple abnormal ECG traces, from where the constraints are computed within the hospital server (before the patient is monitored wirelessly). Then, normal and CVD-affected ECG traces can be fed to our Data Mining (DM) model (i.e., the model that executes on hospital server) to obtain the constraint set (Γi) for all the cases (i.e., normal and all the abnormal cases). Equation (7) shows the constraint set (Γi) for an individual case, where $i = 0, 1, 2, \ldots, N$, N is the maximum number of diseases, and M is the total number of attributes for that particular disease. The number of attributes depends on the attribute selection procedure on each patient. Equation (8) shows the mean values of each attribute (i.e., for all M attributes). On the other hand, Eq. (9) shows the standard deviations against the mean values Ci. Therefore, Eq. (10) demonstrates the valid ranges of each clusters (e.g., normal beat, ventricular tachyarrhythmia, premature ventricular beats, etc.). These constraint set can efficiently be calculated from the proposed DM

$$\Gamma = \{\Gamma 1, \Gamma 2, \Gamma 3, \ldots, \Gamma N\} \tag{7}$$

$$\Gamma \mathrm{i} = \left\{ \mathrm{f}^{\mathrm{i}}_{1},\ \mathrm{f}^{\mathrm{i}}_{2,}\mathrm{f}^{\mathrm{i}}_{3}\ldots\mathrm{f}^{\mathrm{i}}_{\mathrm{M}} \right\} \tag{8}$$

$$\mathrm{Ci} = \left\{ \mathrm{C}^{\mathrm{i}}_{1},\ \mathrm{C}^{\mathrm{i}}_{2,}\mathrm{C}^{\mathrm{i}}_{3}\ldots\mathrm{C}^{\mathrm{i}}_{\mathrm{M}} \right\} \tag{9}$$

$$\mathrm{Pi} = \left\{ \mathrm{P}^{\mathrm{i}}_{1},\ \mathrm{P}^{\mathrm{i}}_{2,}\mathrm{P}^{\mathrm{i}}_{3}\ldots\mathrm{P}^{\mathrm{i}}_{\mathrm{M}} \right\} \tag{10}$$

$$\mathrm{f}^{\mathrm{i}}_{\mathrm{M}=}\mathrm{C}^{\mathrm{i}}\mathrm{M} \pm \mathrm{P}^{\mathrm{i}}_{\mathrm{M}} \tag{11}$$

Patient's mobile can execute a rule-based system following the condition specified by Eq. (11) to detect particular abnormality. During continuous monitoring, the patient's mobile phone can quickly calculate the frequency of characters selected for clustering (i.e., the characters selected by the attribute selection process on the hospital server). s_m is the value of the mth attribute (this value is the frequency count of a particular character from the compressed ECG). F is the set containing the frequency counts of all the characters used for encoding ECG signal. Based on the condition set out by Eq. (11), the compressed ECG packet of

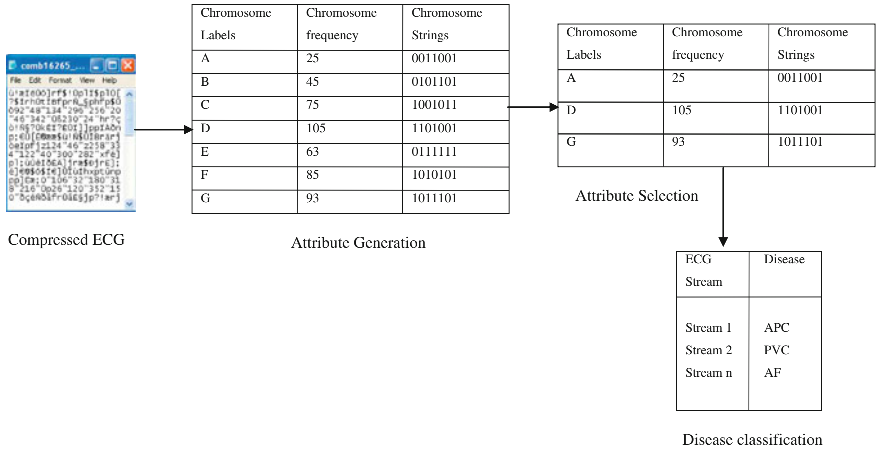


Fig. 3 Disease analysis

the patient can be classified as belonging to a particular class i (where, $i = 1, 2, 3,\ldots,N$).

$$\forall\ s_m \text{in } F : \left(C^i_m - P^i_m\right) \leq s_m \leq \left(C^i_m + P^i s_m\right) \tag{12}$$

3.1.6 Result Analysis

The ECG data are optimized using Genetic Algorithm which helps to take only few data for further processing and efficient clustering is performed to analysis the patients disease. This provides an efficient algorithm for remote health monitoring system (Fig. 3).

4 Conclusion

Patient's mobile phone transmits the compressed ECG packets to the hospital at a convenient time (e.g., may be once a week, when Internet connectivity is available). This model is suitable whenever there is bandwidth restriction on the Internet (i.e., continuous monitoring by the hospital server is not possible). On receipt of the compressed ECG segments, the hospital server calculates the constraints (i.e., Γ values) and forwards it to the patient's mobile phone. The patient's mobile phone then performs continuous monitoring or beat classification. This work could be enhanced as identification of the patient and retrieval of the patients previous records for doctor analysis.

References

1. Palaniappan S, Awang R (2008) Intelligent heart disease prediction system using data mining techniques IEEE/ACS international conference on computer systems and applications. AICCSA, ISBN:978-1-4244-1967-8. pp 108–115
2. Sufi F, Khalil I, Mahmood AN (2011) A clustering based system for instant detection of cardiac abnormalities from compressed ECG. J Expert Syst Appl: Int J 38(5)
3. Mamaghanian H, Khaled N, Atienza D, Vandergheynst P (2011) Compressed sensing for real-time energy-efficient ECG compression on wireless body sensor nodes. IEEE Trans Biomed Eng 58(9):2456–2466
4. Alshamali A, Ghaith T (2009) Combined coding and adaptive thresholding algorithms for ECG compression 5th IEEE transaction on GCC conference and exhibition, pp 1–3
5. Dey N, Dash TP, Dash S (2011) ECG signal denoising by functional link artificial neural network (FLANN). Int J Biomed Eng Technol 7(4):377–389
6. Latha KM, Gouthami S, Ranjith Kumar P (2012) Scoring system approach for assessment of critical illness using mobile phones. Int J Distrib Parallel Syst (IJDPS) 3(1)
7. Nemati E, Deen MJ, Mondal T, McMaster GK (2012) A wireless wearable ECG sensor for long-term applications. IEEE Commun Mag
8. Mehta SS, Lingayat NS Comparative study of QRS detection in single lead and 12-lead ECG based on entropy and combined entropy criteria using support vector machine. J Theor Appl Inf Technol
9. Kumar V, Saxena SC, Giri VK (2005) Dilbag singh, improved modified AZTEC technique for ECG data compression: effect of length of parabolic filter on reconstructed signal. Comput Electri Eng 334–344
10. Nygaard R, Melnikov G, Katsaggelos AK (2001) *Fellow* a rate distortion optimal ECG coding algorithm. IEEE Trans Biomed Eng 48(1):28–40
11. García J, Martínez I, Sörnmo L, Olmos S, Mur A, Laguna P (2002) Remote processing server for ECG-based clinical diagnosis support. IEEE Trans Inf Technol Biomed 6(4)
12. Chen D, Yang S (2008) Compression of ECG signal using video codec technology-like scheme. J Biomed Sci Eng 1(1)
13. Zhang Y, Xiao H (2009) Bluetooth-based sensor networks for remotely monitoring the physiological signals of a patient. IEEE Trans Inf Technol Biomed 13(6)
14. Oresko JJ (2010) Real-time cardiovascular disease detection via electrocardiogram. Process IEEE Trans Inf Technol Biomed 14(3)

CMOS 2nd Order Gm-C Intermediate Frequency Band Pass Filters for Wireless Systems

P. Sampath, R. Harikumar and K. Gunavathi

Abstract In this paper the design and implementation of CMOS Gm-C 2nd order band pass filters operating at center frequency 45 MHz with a bandwidth of 3 MHz and Quality factor (Q) of 15 are presented. The Band Pass Filter (BPF) is a Biquad structure with Operational Transconductance Amplifier (OTA) as the basic building block and is implemented in Cadence Analog Design Environment using CMOS 0.18 μm technology. The performance of the BPF designed with different OTA's are compared in terms of Q, voltage gain, IIP3, 1-dB compression and noise parameters. These filters can be operated at center frequency of 45 MHz under 1.8 V power supply. The filters are suitable for intermediate frequency (IF) range in the transmitter of most of the wireless systems operating at 900 MHz and 2.4 GHz.

Keywords Biquad · Band pass filter · OTA · Wilson current mirror · Cascode current mirror

P. Sampath (✉) · R. Harikumar
Department of Electronics and Communication Engineering, Bannari Amman Institute of Technology, Sathyamangalam, India
e-mail: sampathpalaniswami@gmail.com

R. Harikumar
e-mail: harikumarrajaguru@gmail.com

K. Gunavathi
Department of Electronics and Communication Engineering, PSG College of Technology, Coimbatore, India
e-mail: kgunavathi2000@yahoo.com

Mohan S. and S. S. Kumar (eds.), *Proceedings of the Fourth International Conference on Signal and Image Processing 2012 (ICSIP 2012)*, Lecture Notes in Electrical Engineering 222,
DOI: 10.1007/978-81-322-1000-9_22,

1 Introduction

The rapidly increasing use of wireless devices has accelerated the need for implementation of high speed, low cost on-chip IF filters. Realization of on-chip band pass filters for narrow band selection is challenging task in wireless receivers. IF band pass filters are needed for the channel selection and filtering. In super-heterodyne receivers proper filtering is mandatory and is done by external surface acoustic wave (SAW) filters. Filtering can be done by making use of active filters. The Gm-C circuits represent a popular technique of integrated realization of high frequency continuous time filters [1]. 6th order Chebychev type-1 BPF operating at center frequency of 46 MHz with bandwidth of 2.046 MHz is represented using Biquad Gm-C structure [2]. This filter uses negative resistance cells for high Q and the current consumption is 8.87 mA. High-Q, high linear micro electro mechanical filter with center frequency of 71 MHz is proposed with IC compatibility [3]. In this filter the termination resistances are in the order of KΩ and MEMS based design. A Gm-C frequency-tunable RF bandpass filter with multigated transistors operating in the band of 45–470 MHz is presented [4]. This filter is implemented in 65 nm technology and hence the power consumption is low but tuning is difficult for narrow bandwidth. A 10.7-MHz fully balanced, high-Q, wide-dynamic-range current-tunable Gm-C band pass filter is presented [5]. This filter uses adder and low –Q based BPF. The Q of this filter is very low. The Gm-C filter offers many advantages in terms of low-power and high frequency capability. Gm-C filters can operate in a wide range of frequencies from several hundred of KHz to more than 100 MHz. The Q of Gm-C filters can be adjusted by controlling the output impedance even at lower frequencies. In this paper 2nd order Gm-C BPF's to operate at center frequency of 45 MHz and bandwidth 3 MHz using Biquad structure are presented. OTA is used as building block in the Biquad structure. Three different OTA's are used to implement the BPF and their performance is compared.

The working of different OTA's and their performance is discussed in Sect. 2. The construction of 2nd order BPF with different OTA's and the simulation results of the 2nd order BPF with the three OTA's are given in Sect. 3 and the results are discussed. Finally the conclusion is drawn in Sect. 4.

2 Operation of OTA's

The operational transconductance amplifier (OTA) is used as basic building block in many switched capacitor filters. OTA is basically an op-amp without an output buffer and can only drive capacitive loads [6, 7]. An OTA is an amplifier where all nodes are low impedance except the input and output nodes. A useful feature of OTA is that its transconductance can be adjusted by the bias current. Filters made using the OTA can be tuned by changing the bias current I_{bias}. Two practical concerns when designing an OTA for filter applications are the input signal amplitude and the parasitic input and output capacitances. The external capacitance should be large compared to the input or output parasitic capacitance of the OTA. Large signals cause the OTA gain to become non-linear. This limits the maximum frequency of a filter built with an OTA and causes amplitude or phase errors [8]. These errors can be minimized with proper selection of $I_{bias.}$ The transconductance of the OTA is given by

$$g_m = \frac{i_{out}}{v^+ - v^-} \tag{1}$$

where, i_{out} is the output current of the OTA and v+ and v- is the differential input voltage to the OTA. The voltage gain of the OTA is given by.

$$A_v = \frac{v_{out}}{v^+ - v^-} \tag{2}$$

where, v_{out} is the output voltage of the OTA. The performance of simple OTA is limited by its input and output voltage swing. To overcome the limits of simple OTA and improve the performance, OTA with Wilson Current Mirror (WCM-OTA) and OTA with Cascode Current Mirror (CCM-OTA) are used [8]. The WCM-OTA has a low output voltage swing and the performance can be further improved by using a Cascode current mirror. The OTA's used in designing the band pass filter [6] are shown in Fig. 1.

The circuit parameters for OTA's used are given in Tables 1 and 2.

The transient analysis is performed for all the three OTA's to find their transconductance and the simulation gives a transconductance of 160, 52.44 and 58.68 μS for Simple OTA, WCM-OTA and CCM-OTA respectively.

3 Biquad BPF and Simulation Results

The Biquad implement's the band pass function [7, 9]. The Biquad implemented for filter design is shown in Fig. 2.

The input condition for the Biquad structure to make it act as BPF is $\mathbf{V_2 = V_{in}}$, $\mathbf{V_1 =}$ **ground and** $\mathbf{V_3 =}$ **ground**, where V_1, V_2, V_3 are the inputs to the Biquad as shown in Fig. 2. Both the OTA's in the Biquad are the same and hence the transconductance of each stage is also same. The transfer function of the second order band pass filter is given by

$$H(s) = \frac{sC_1 g_m}{s^2 C_1 C_2 + sC_1 C_2 + g_m^2} \tag{3}$$

The specifications of the proposed filter design are given in Table 3.

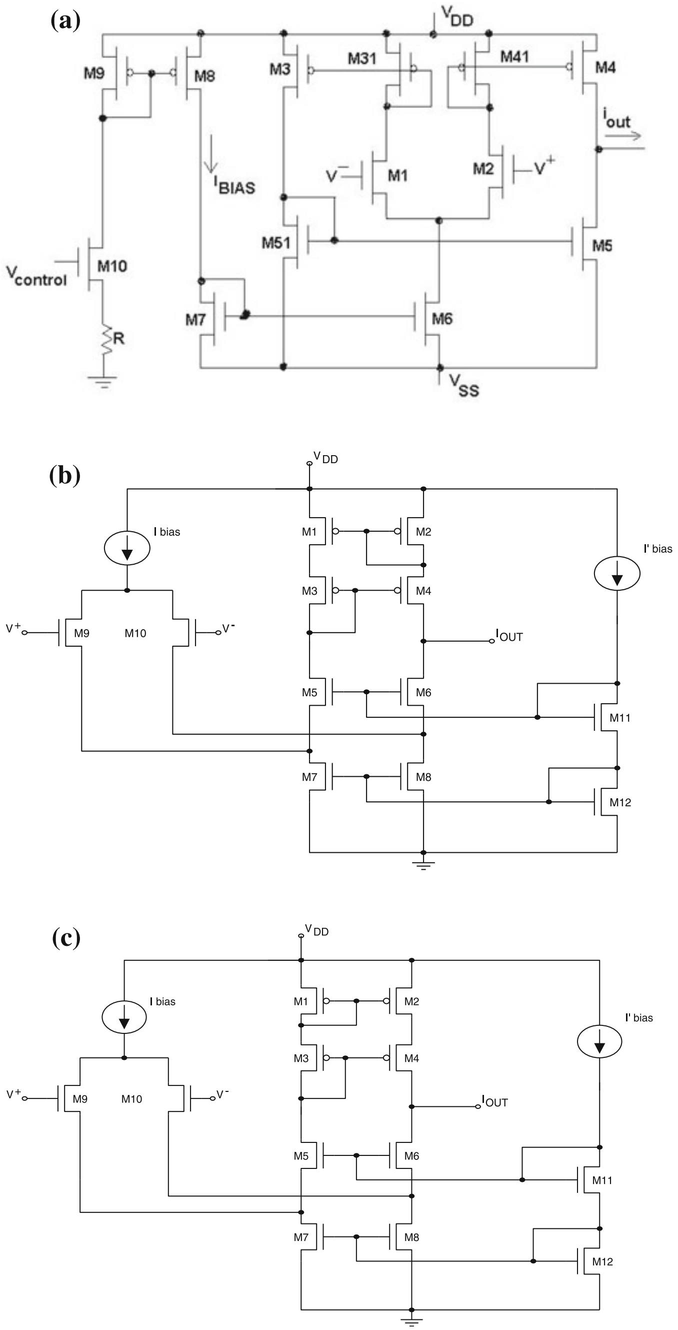


Fig. 1 Circuit of different OTA's used **a** Simple OTA circuit **b** WCM-OTA circuit **c** CCM-OTA circuit

Table 1 Circuit parameters of the simple OTA

Parameters	Simple OTA
M3, M4, M8, M9, M41 and M31(W/L)	22/0.18 μm
M1,M2, M51, M5, M6, M7 and M10(W/L)	2/0.18 μm
R	100 KΩ

Table 2 Circuit parameters of WCM-OTA and CCM-OTA

Parameters	WCM-OTA	CCM-OTA
M1, M2, M3 and M4 (W/L)	9/0.18 μm	9/0.18 μm
M5, M6, M7, M8, M11 and M12 (W/L)	3.08/0.18 μm	3.08/0.18 μm
M9 and M10	10/0.18 μm	8/0.18 μm

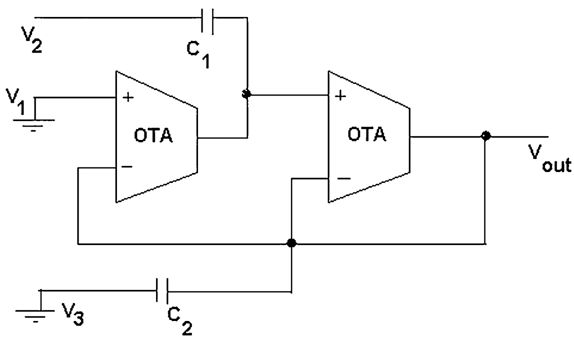


Fig. 2 Circuit of Biquad 2nd order BPF

Table 3 Specifications of the second order BPF

Application	Mobile applications
Filter type	Active Gm-C
Center frequency	45 MHz
Bandwidth	3 MHz
Order of the filter	2
Technology	0.18 μ CMOS
Power supply	1.8 V

The ac simulation is performed with the three OTA's individually for the BPF, to find the gain and bandwidth of the filter. The ac response of the filter for the OTA's is given in Fig. 3.

The ac response of the 2nd order BPF with different OTA's shows that the filter with CCM-OTA gives more gain and Bandwidth, the filter with simple OTA gives relatively large bandwidth and the filter with Simple-OTA give a Q near to filter with WCM-OTA and more bandwidth than WCM-OTA. The ac response of the filter with different OTA's is given in Table 4.

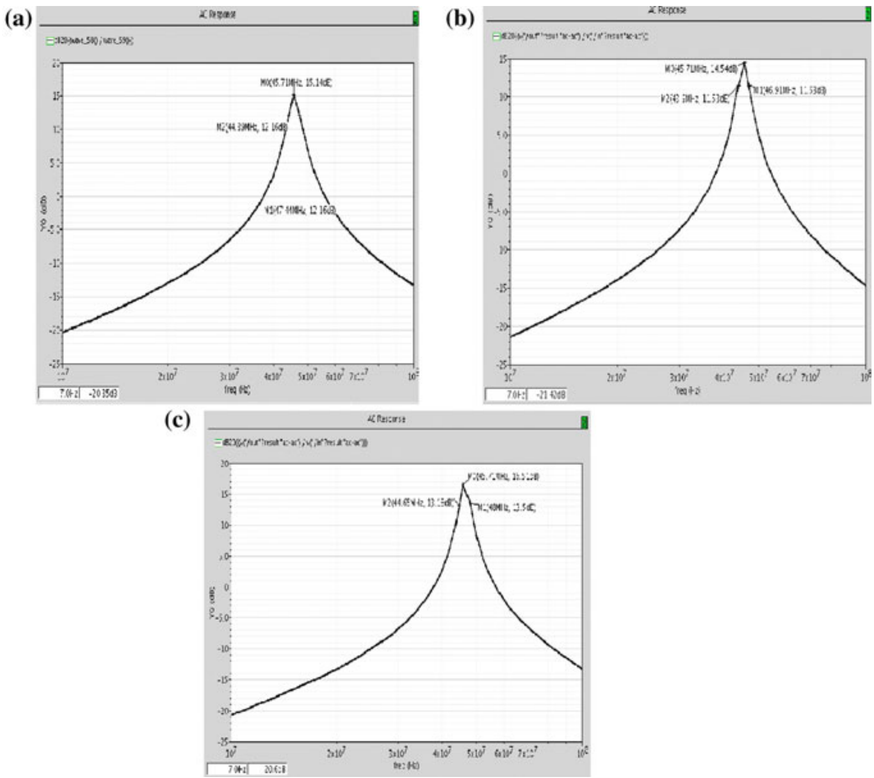


Fig. 3 AC response of the filter with different OTA's **a** Simple OTA **b** WCM-OTA **c** CCM-OTA

Table 4 AC response of the filter with different OTA's

Filter with type of OTA	Design values	Simple OTA	WCM-OTA	CCM-OTA
Center frequency	45 MHz	45.71 MHz	45.71 MHz	45.71 MHz
Gain	>10 dB	15.14 dB	14.54 dB	16.51 dB
Bandwidth	3 MHz	3.05 MHz	3.01 MHz	3.35 MHz
Q	15	14.98	15.18	13.64

The S-parameter analysis is performed to find the S-parameters, Insertion loss, Power gain and available gain of the circuit and stability of the circuit. The S-parameter simulation results of the filter for different OTA's is given in the Table 5.

The S-parameter simulation of the filter shows that the input and output impedance matching is good for the designed filter with different OTA's since the values of S21and S12 is high and values of S11 and S22 are low. The BPF with WCM-OTA and CCM-OTA provide maximum power gain than the Simple-OTA

Table 5 S-parameters of the filter with different OTA's

S-parameters	Simple OTA	WCM-OTA	CCM-OTA
S_{21}	−3.184 dB	−1.36 dB	−1.15 dB
S_{11}	−16.39 dB	−14.39 dB	−19.98 dB
S_{12}	−3.607 dB	−1.36 dB	1.118 dB
S_{22}	−7.777 dB	−16.11 dB	−12.62 dB
Power gain G_P	−3.084 dB	14.37 dB	12.45 dB
Transducer gain G_T	−3.184 dB	−1.36 dB	−1.118 dB
Available gain G_A	−2.935 dB	−1.253 dB	−0.88 dB
Bif (Δ)	0.6993	0.4922	0.3486
Rollet stability factor (K)	1.059	1.002	1.001

Table 6 PSS analysis of the filter with different OTA's

PSS parameters	Simple OTA	WCM-OTA	CCM-OTA
1-dB compression point	−29.593 dBm	−38.635 dBm	−38.609 dBm
IIP3	−29.94	−29.6825	−29.6957
Voltage gain	−17.07 dB	−29.88 dB	−29.83 dB
Harmonic distortion	−24.2 dB	−36.1 dB	−36.06 dB

Table 7 PN Analysis of the filter with different OTA's

PN parameters	Simple OTA	WCM-OTA	CCM-OTA
Input noise	−149.5 dB	−141 dB	−141.5 dB
Output noise	−152.7 dB	−142.6 dB	−142.7 dB

BPF. The stability factor Bif (Δ) of less than one and K greater than one shows that all the three BPF's are stable.

The Periodic Steady State (PSS) analysis is performed for the BPF to find its 1 dB compression point, 3rd order Input Intercept Point (IIP3), Voltage gain and harmonic distortion. The parameters of PSS analysis are given in Table 6.

IIP3 is measured using a two-tone test, where the two input tones are the first order tones. The IIP3 is defined as the cross point of the power for the 1st order tones and the power for the 3rd order tones. The Voltage gain is found as a ratio of the IF Output to the IF input in dB. The Harmonic distortion is characterized as the ratio of the power of fundamental signal to the sum of power at the harmonics.

The Periodic noise (PN) analysis is performed for the BPF with the three OTA's to visualize the contribution of different noise sources in the total noise. This noise analysis gives the noise performance of the device which contributes the maximum noise. The input and output noise of the filter with the OTA's is given in the Table 7.

4 Conclusion

The design and implementation of 2nd order CMOS band pass filters using OTA as the basic element in the Biquad structure has been presented and is shown from the simulation results that the filters can work at a low supply voltage of 1.8 V. The filters were designed for center frequency of 45 MHz and bandwidth of 3 MHz with a Q of 15. The simulation results of the filters show that the Q is high for BPF with WCM-OTA than the BPF with CCM-OTA and the BPF with simple OTA. The S-parameter analysis also shows that the BPF with WCM-OTA and CCM-OTA perform well in terms of matching and gain at the designed frequency and bandwidth. The PSS simulation shows that the distortion is less in BPF with WCM-OTA and CCM-OTA. The input and output noise is low for BPF with Simple OTA. The BPF consumes a power of 117 μW with simple OTA, 55 μW with WCM-OTA and 75 μW with CCM-OTA. All three filter structures can be used in IF channel selection and filtering in the transmitter of wireless applications as they consume very less power and provide good filtering characteristics.

References

1. Qureshi MS, Allen PE (2005) 70 MHz CMOS Gm-C IF filter. IEEE J Solid State Circuits 6:46–49
2. duan C, Li W (2011) A 46 MHz Biquad Gm-C high Q bandpass filter design for wireless application. DASC ‘11 proceedings of the 2011 IEEE ninth international conference on dependable, autonomic and secure computing, pp 69–72
3. Babazadeh F, Keshmiri SH (2007) Design and simulation of a High-Q, low motional resistance, highly linear if micro electro mechanical filter. International conference on perspective technologies and methods in MEMS design, 2007. MEMSTECH
4. Jolivet S, Amiot S, Lohy D, Jarry B, Lintignat J (2011) A 45–470 MHz Gm-C tunable RF bandpass filter for TV tuners designed in 65 nm CMOS. In: Proceedings 9th International Conference on New Circuits and Systems Conference, pp 13–16
5. Worawat Sa-Ngiamvibool1 and Banlue Srisuchinwong2, 10.7-MHz fully balanced, high-Q, wide-dynamic-range current-tunable Gm-C bandpass filter. ECTI Trans Electr Eng Electr Commun 5(2):13–22
6. Johns DA, Martin K (1997) Analog integrated circuit design. Wiley, New York
7. Baker RJ, Li HW, Royce DE (1998) CMOS-circuit design, layout and simulation. IEEE Press, London
8. Daoud H, Salem SB, Zouari S, Loulou M (2006) Folded cascode OTA design for wide band applications. Design and test of integrated systems in nanoscale technology
9. Geiger RL, Sánchez-Sinencio E (1985) Active filter design using operational transconductance amplifiers: a tutorial. IEEE Circuits Devices Mag 1:20–32

Design of Two Element and Four Element Printed Dipole Array Antennas for Wireless Communication Applications

Chenniappan Poongodi, Arumugam Shanmugam and P. Prabhu

Abstract In Multiple Input Multiple Output (MIMO) system, both transmitter and receiver are provided with more than one antenna to increase the channel capacity and quality. This paper describes the simulated results of microstrip balun fed two and four element printed dipole array antenna using ADS software for wireless communication applications and revealed that the return loss of 4 element side by side array antenna is −24.04 dB at 5.6 GHz. When the spacing between the elements is less than 0.5λ, the effect of mutual coupling is more. The mutual impedance between antenna elements is computed from the measured S parameters and it is below 24 dB for the measured frequency range 4–6.8 GHz. The designed antennas are characterized by measuring return loss, directivity, radiation pattern and gain. Gain and directivity of four element array is 8.85 dBi and it radiates 18.9 mW power. The performance of arrays is evaluated using capacity as the metric based on Monte Carlo realizations of the channel with inclusive of mutual coupling between dipoles.

C. Poongodi (✉) · A. Shanmugam · P. Prabhu
Bannari Amman Institute of Technology, Sathyamangalam, Tamilnadu, India
e-mail: poongi_cj@yahoo.co.in

A. Shanmugam
e-mail: dras_bit@yahoo.com

P. Prabhu
e-mail: prabhu.beece66@gmail.com

Mohan S. and S. S. Kumar (eds.), *Proceedings of the Fourth International Conference on Signal and Image Processing 2012 (ICSIP 2012)*, Lecture Notes in Electrical Engineering 222,
DOI: 10.1007/978-81-322-1000-9_23,

1 Introduction

Multiple Input Multiple Output (MIMO) wireless systems have demonstrated the potential to increase communication spectral efficiency in a rich multipath environment [1]. The demand for spectral efficiency in wireless communication is ever increasing. It has been established in [1] that using antenna arrays at the transmitter and the receiver can increase the capacity and it is proportional to minimum of number of transmitting elements and number of receiving elements. The capacity depends mainly on the channel and the antenna characteristics. The capacity can be improved by proper design of antenna elements [2] and choosing appropriate array configuration [3]. Multiple-Input Multiple-Output (MIMO) wireless systems have demonstrated the potential to increase communication spectral efficiency in a rich multipath environment. From an antennas perspective, different array configurations and types of element have been proposed and analyzed for MIMO links. Therefore it is important to know how various array configurations are performing in the case of MIMO system. In both military and commercial applications there have been ever growing demands for antennas. In this article printed dipole array is presented for potential use in wireless communication. The dipoles are fed through a microstrip balun. Advantages of the printed dipole antennas are low cost, compact size, ease of fabrication and low profile [4]. The mutual coupling [5] also changes the capacity considerably. When the spacing between the elements is less than 0.5λ, the effect of mutual coupling is more. The aim of this work is to design and evaluate the performance of the two element and four element side by side array configurations for Wireless local area networks (IEEE 802.11) applications.

2 Antenna Design

The antenna is designed on a substrate of thickness, h = 1.4 mm with dielectric constant, $\epsilon r = 4.5$. A dipole antenna usually needs a balanced feed for practical operation.

The electric field of microstrip lines is mainly normal to the substrate however the electric field across the gap between the arms of the dipole is along its length, thus, the dipole cannot be fed directly from a microstrip line. This requires alternative feeding mechanisms, for example co planar strips or microstrip to slot line cross junction. We chose to excite the dipole with a printed balun [6]. A balun is a device used to balance an unbalanced transmission line. The printed dipole with the integrated balun features a broadband performance [7] and has found applications in wireless communications [8] and antenna arrays [9]. For simulation the length of the designed antenna, L = 35 mm and width, W = 55 mm, space between dipole arm is g = 1 mm. Figure 1 shows geometry of a printed dipole antenna with adjusted integrated balun. The geometry size of dipole arm and balun is shown in Table 1. In MIMO systems multiple printed dipole antennas are used. The two element printed dipole antenna and four element array antennas are shown in Figs. 2 and 3 respectively.

3 MIMO System Model

In MIMO system, the dipole antennas are placed side-by-side within a finite length to form a uniform linear array. The length of each dipole is assumed to be the odd multiplies of half wavelength. First assume all N_T transmit antennas and N_R receive antennas will be powered up. The receive antennas are equally spaced within a limited length L_R, which is equal to several wavelengths for typical consumer wireless equipment. The complex envelope of the received signal at the antenna array after matched filtering is given by

$$y = Hx + n \tag{1}$$

where x is the transmit vector, y is the receive vector, H is the $N_R \times N_T$ channel matrix, and n is the additive white Gaussian noise (AWGN) vector at a given instant in time. Throughout the paper, it is assumed that the channel matrix is random and that the receiver has perfect channel knowledge [10]. It is also assumed that the channel is memoryless, i.e., for each use of the channel an independent realization of H is drawn. A general entry of the channel matrix is denoted by $\{h_{ij}\}$. This represents the complex gain of the channel between the jth transmitter and the ith receiver. With a MIMO systems consisting of NT transmit antennas and N_R receive antennas, the channel matrix is written as

$$H = \begin{bmatrix} h_{1,1} & \ldots & h_{1,n_T} \\ .. & .. & .. \\ h_{n_R,1} & \ldots & h_{n_R,n_T} \end{bmatrix} \tag{2}$$

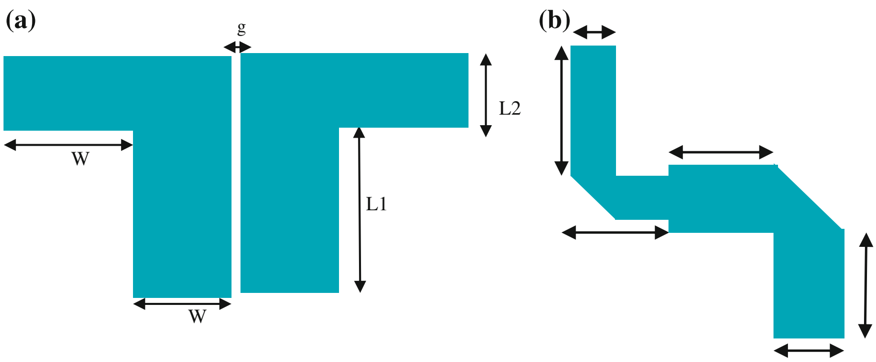


Fig. 1 **a** Configuration of the printed dipole antenna-radiating element, **b** configuration of the Balun type feed for dipole antenna

Table 1 The designed geometry size of printed dipole antenna and balun

Dipole arm		Balun	
Parameter	Value	Parameter	Value (mm)
L1	23 mm	l1	7.6 mm
L2	12 mm	l2	7 mm
W1	11 mm	w1	2.8 mm
W2	14 mm	w2	5 mm
g	1 mm	w3	4 mm
Overall dimension	35*55 mm^2	w4	1.6 mm

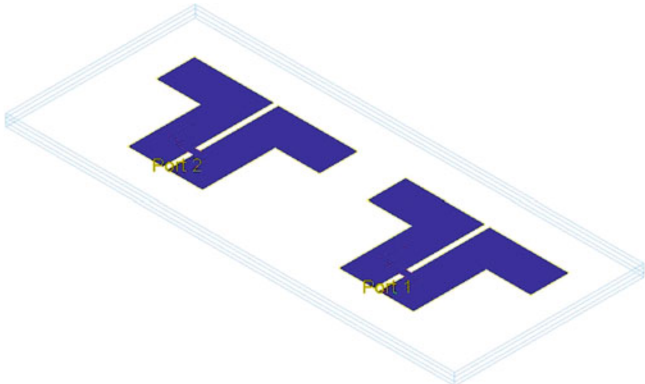


Fig. 2 Two element printed dipole array antenna on FR5 substrate

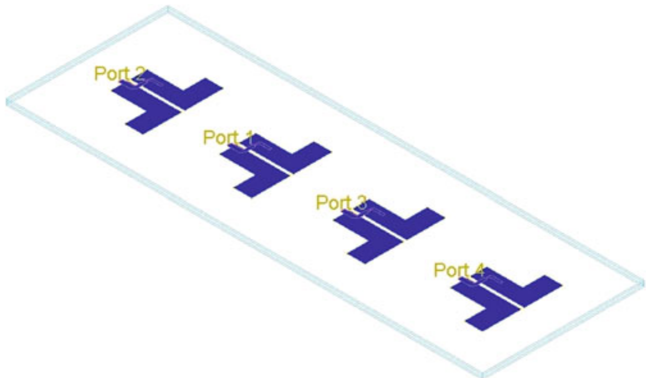


Fig. 3 Four element printed dipole antenna in side by side array configuration

In a rich scattering environment with no line of sight (LOS), the channel gains are usually independent identically distributed Gaussian random variables.

3.1 Mutual Coupling

The principal function of an antenna is to convert an electromagnetic field into an induced voltage or current to be measured. However, the measured voltage at each antenna element will depend not only on the incident field, but also on the voltage on each other elements. Essentially, the received voltage on each element will induce a current on the element which in turn radiates a field which affects the surrounding element, i.e. the elements are said to be mutually coupled. Mutual coupling is well known in the antenna community, since coupling between antenna elements is one of the most important properties to consider in an antenna design. So to include the coupling effect in the model for the received signal Eq. (1) is modified to

$$y = CHx + n. \tag{3}$$

It is then natural to include the coupling effects into the channel by combining the two terms into a new channel matrix H_{mc}. Eq. (3) includes coupling at the receiving antenna elements, but several closely spaced elements at the transmitter will also experience mutual coupling [11]. Thus, including this effect at both transmitter and receiver, the expression for the channel becomes

$$H_{mc} = C_b H\, C_m. \tag{4}$$

where the coupling matrix at the base C_b is $N_R \times N_T$ and the corresponding matrix at the mobile C_m is $N_T\, N_T$. Using fundamental electromagnetic and circuit theory, the coupling matrix of an array element is derived from the simulated S parameter values.

3.2 SISO Channel Capacity

The ergodic (mean) capacity of a SISO system ($N_T = N_R = 1$) with a random complex channel gain h_{11} is given by [5]

$$C = E_H\{\log_2(1 + \rho|h_{11}|^2\}. \tag{5}$$

where ρ is the average signal to noise (SNR) ratio at the receiver branch. If h_{11} follows a Chi squared distribution with two degrees of freedom [11] Eq. (5) can be written as [5]

$$C = E_H\{\log_2(1 + \rho\chi_2^2\}. \tag{6}$$

where χ_2^2 is a Chi squared distributed random variable with two degrees of freedom.

3.3 MIMO Channel Capacity

For a narrow band MIMO channel, when CSI is not known at the transmitter, the capacity is given by [5]

$$C = E\ [\log_2(\det(I_{Nr} + \rho H\ H^+ /\ N))] \tag{7}$$

where ρ is the average signal to noise ratio at each receiver and H is the $N_R \times N_T$ channel matrix. $N = min(N_T,\ N_R)$. This expression assumes that the available transmit power ρ is uniformly allocated to the N_T transmit antennas, which is the practical approach when the transmitter has no knowledge of the channel. It is easily realized from this expression that a large capacity hinges on the presence of a rich scattering environment, being directly related to the rank of the channel matrix. Conversely, little or no scattering will result in a channel matrix of unit rank and thus low capacity. More often each element of H is taken to be i.i.d complex Gaussian distributed random variable signifying that each pair of transmit and receive antennas experiences independent fading. But, this is not true in practical situation. Because of spacing and mutual coupling between the elements, independent fading is not a valid assumption. In [5], the effect of coupling between antenna elements is included in the channel matrix. Now the capacity is modified as

$$C = E\ [\log_2(\det(I_{Nr} + \rho H_{mc} H_{mc}^+ /\ N)^)]. \tag{8}$$

The capacity is computed for a large number of channel realizations.

4 Result and Discussion

The simulation is carried out by the method of moment’s technique (ADS software) [12]. The radiation pattern of the four element printed dipole array antennas at 5.6 GHz are shown in Fig. 4, which shows omnidirectional radiation pattern.

Figure 5 shows the simulated return loss of the two element printed dipole antenna is −27.7 dB at 5.6 GHz and Fig. 6 shows the return loss of the four element printed dipole antenna is −24 dB at 5.6 GHz. Since this array antennas are having minimum return loss at 5.6 GHz, which might be used for wireless communications. Gain and directivity of the two element array is shown in Fig. 7, which provides 5.92 dBi gain and omnidirectional radiation pattern. Two element

Fig. 4 Radiation pattern of four element printed dipole antenna

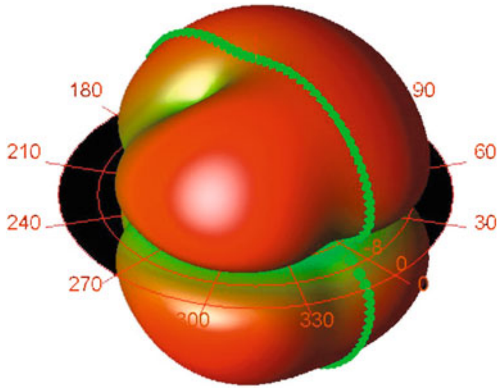


Fig. 5 Simulated return loss of the two element dipole printed antenna at 5.6 GHz

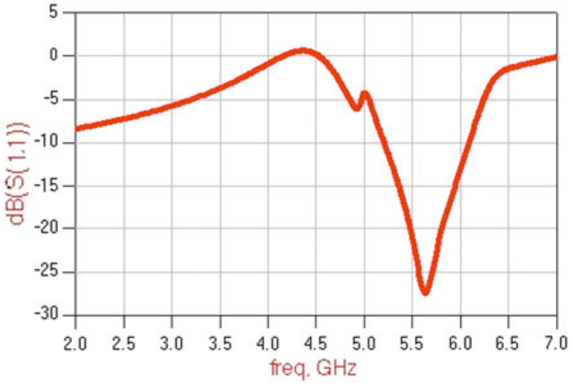
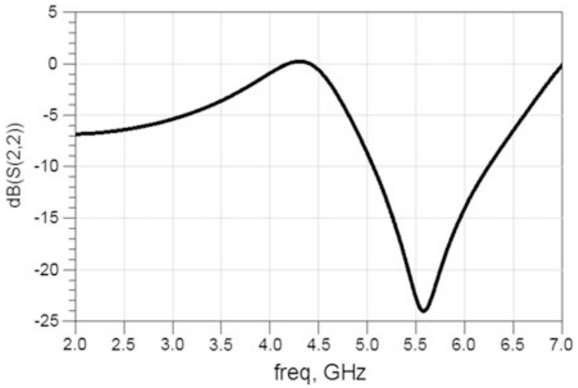


Fig. 6 Simulated return loss of the four element printed dipole array antenna at 5.6 GHz



arrays provides better gain compare to single antenna system at 5.6 GHz and this gain increases with number of antenna elements. The circularly polarized electric field pattern of the four element array is shown in Fig. 8. Absolute electric field

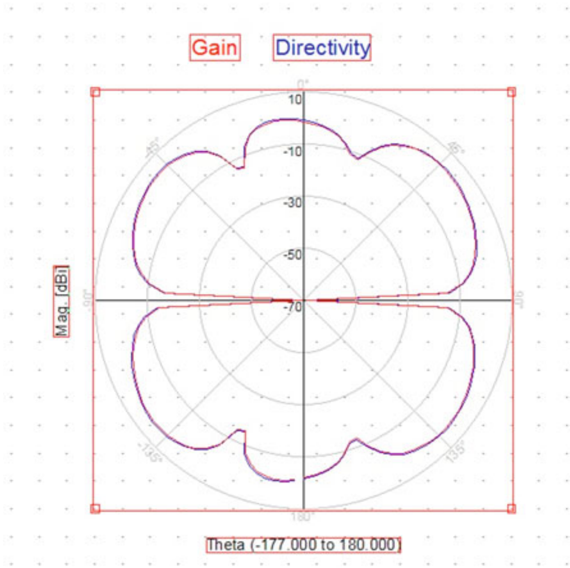


Fig. 7 Simulated gain and directivity of two element printed antenna

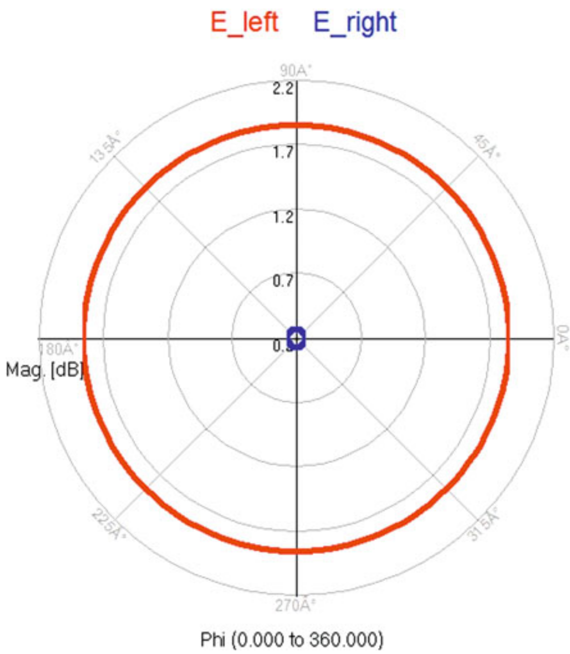


Fig. 8 Circularly polarized electric field pattern of four element array antenna

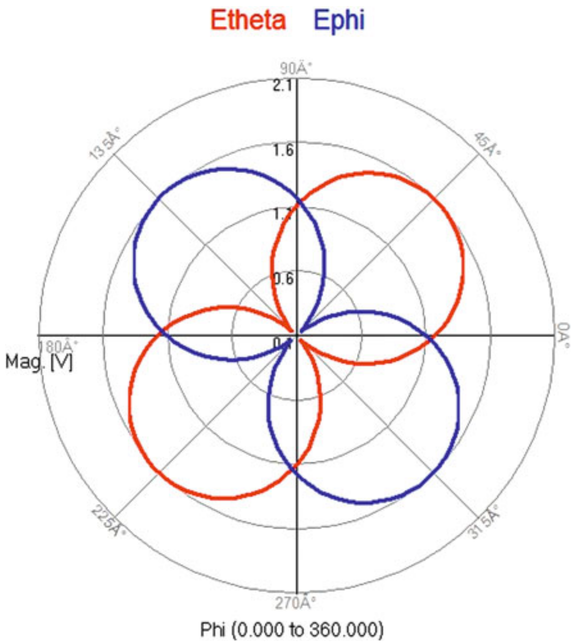


Fig. 9 Absolute fields of four element array antenna

pattern of the array antenna is shown in Fig. 9. Here E_θ and E_ϕ field components are mutual perpendicular to each other and produces the figure eight pattern.

The gain, directivity, radiated power; maximum intensity and efficiency of different antenna configurations are shown in Table 2. Four element array antennas provide better results in terms of gain, directivity, radiated power and also maximum intensity. Figure 10 compares ergodic capacity of single input single output antenna with multiple 4×4 array type printed dipole antenna with and without mutual coupling. The receiver signal-to-noise ratio (SNR) varied from 3 to 12 dB and 10^4 Monte Carlo runs were performed to compute the average capacity.

The mutual coupling between antenna elements is computed from the measured S parameters. Array antenna elements of size four provides better performance than single antenna system and also mutual coupling between the array elements, decrease the channel capacity compare to without coupling array antennas. Thus the use of multiple antennas greatly improves the spectral efficiency in communication system.

Table 2 Simulated results of printed dipole antenna in different configuration

Antenna type	Gain (dBi)	Directivity (dBi)	Radiated power (mW)	Max intensity (mW/ster.)	η(%)
Single printed dipole	3.56	3.57	1.20	1.112	89
2 element array	5.92	5.92	4.38	1.514	98
4 element array	8.85	8.85	18.9	11.39	98

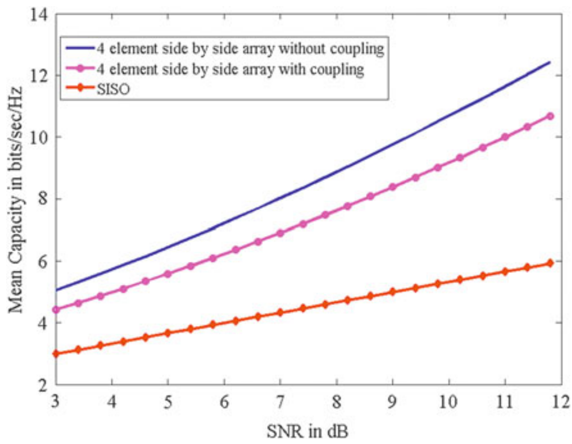


Fig. 10 Mean capacity as a function of SNR for SISO and MIMO configuration

5 Conclusion

The array of printed dipole with an adjusted integrated balun is developed. The antenna has small size and easy to integrate with circuit on the same dielectric, resulting in the reduction of fabrication cost and required volume of whole system. With a help of printed antenna, the analysis of two and four element array antenna noticed that the array antenna produces an omnidirectional radiation pattern and seems to be a good for wireless applications. Finally the capacity of 4×4 array with and without mutual coupling is evaluated using Monte Carlo simulations and compared with single input single output antenna system.

References

1. Foschini BGJ, Gans MJ (1998) On limits of wireless communications in a fading environment when using multiple antennas. Wireless Personal Comm 6: 311–335 (Mar)
2. Waldschmidt C, Fugen T, Wiesbeck W (2002) Spiral and dipole antennas for indoor MIMO-systems. IEEE Antennas Wireless Propagat Lett 1(1):176–178
3. Martin CC, Winters JH, Sollenberger NR (2001) MIMO radio channel measurements: performance comparison of antenna configurations. In Proceeding IEEE 54th Veh. Technol. Conf., vol.2, Atlantic City, NJ, Oct. 7–11, (2001), 1225–1229
4. Li X, Yang L, Gong S-X, Yang Y-J (2009) Dual-band and wideband design of a printed dipole antenna integrated with dual band balun. Prog Electromagnet Res Lett 6:165–174
5. Svantesson T, Ranheim A (2001) Mutual coupling effects on the capacity of multielement antenna systems. In: Proceeding IEEE international conference acoustics, speech, and signal processing, vol 4, Salt Lake City, 7–11 May, pp 2485–2488
6. Michailidis E, Tsimenidis C, Chester G (2007) Printed dipole array for MIMO wireless networks. In: Proceeding IEEE Loughborough antennas and propagation conference, April 2007, pp 117–120
7. Edward B, Rees D (1987) A broadband printed dipole with integrated balun. Microwave J, May, pp 339–344

8. Chuang H-R, Kuo L-C (2003) 3-D FDTD design analysis of a 2.4 GHz polarization-diversity printed dipole antenna with integrated balun and polarization-switching circuit for WLAN and wireless communication applications. IEEE Trans Microwave Theor Tech 51(2): 374–381
9. Bayard J-PR (1994) Analysis of infinite arrays of microstrip-fed dipoles printed on protruding dielectric substrates and covered with a dielectric radome. IEEE Trans Antennas Propag 42(1): 82–89
10. Shuangqing W, Goeckel DL, Janaswamy R (2003) On the asymptotic capacity of MIMO systems with fixed length linear antenna arrays. In: IEEE International Conference on communications, vol 4, Amherst, MA, 11–14 May 2003, pp 2633–2637
11. Svantesson T (2001) A physical MIMO radio channel model for multi-element multi-polarized antenna systems. In: Proceedings IEEE 54th vehicular technology conference, vol 2, Atlantic City, NJ, 7–11 Oct 2001, pp 1083–1087
12. Advanced Design System (ADS) (2011) Agilent Technologies, USA

Multiple Image Encryption Using Phase Manipulation and SCAN Methods

H. T. Panduranga and S. K. Naveen Kumar

Abstract In this paper we propose a new concept of combined image encryption using phase manipulation and SCAN methods. Entire encryption process involves two stages where multiple images to be encrypted are applied to phase manipulation block. In first stage Fourier Transform (FT) is applied to get phase and magnitude of all input images. Phase of all images are scrambled to get modified image after applying Inverse Fourier Transform. In second stage these modified images are scrambled by using SCAN method. SCAN method finally gives resultant encrypted image by rearranging the pixels positions of modified image. Experiment is conducted for dual and multiple images using MATLAB software. From the experiment we obtained highly scrambled image at the end of encryption process. Decryption process employs exactly reverse process of encryption which results in the reconstructed images.

Keywords Phase manipulation · SCAN patterns · Encryption · Decryption

1 Introduction

Today's word totally dependent on information and communication technology. Due to fast growth in communication technology there is huge demand for internet to exchange the information and hence there is a necessity to give security to all

H. T. Panduranga (✉) · S. K. Naveen Kumar
Deptartment of Studies in Electronics, Hemagangothri PG Center Hassan,
University of Mysore, Mysore, Karnataka, India
e-mail: ht_pandu@yahoo.co.in

S. K. Naveen Kumar
e-mail: nave12@gmail.com

Mohan S. and S. S. Kumar (eds.), *Proceedings of the Fourth International Conference on Signal and Image Processing 2012 (ICSIP 2012)*, Lecture Notes in Electrical Engineering 222,
DOI: 10.1007/978-81-322-1000-9_24,

these information. Image is also a one of the most important data which carries more information and require security. In this regard we are developing a new concept for image encryption using phase manipulation and SCAN methods. From the literature survey we come to know that many researchers are working in the field of image encryption in their own methods. All the methods for image encryption will have their own identity, strength and weakness.

Tzung-Her Chen [1] (n + 1, n + 1) visual secret sharing (VSS) is used to encode (encrypt) a secret image into n meaningless share images to be superimposed later to decode (decrypt) the original secret by human visual system after collecting all n + 1 secret images. Prasanna et al. [2] have presented an image encryption method with magnitude and phase manipulation using carrier images. Here they used the concept of carrier images and one dimensional Discrete Fourier Transform for encryption purpose and it deals with private key cryptosystem, works in the frequency domain. Maniccam and Bourbakis [3] have presented a new methodology which performs both lossless compression and encryption of binary and gray-scale images. The compression and encryption schemes are based on SCAN patterns generated by the SCAN methodology. The SCAN is a formal language-based two-dimensional spatial- accessing methodology which can efficiently specify and generate a wide range of scanning paths or space filling curves. Kachris et al. [4] are implemented SCAN algorithm in FPGA to know the performance of SCAN algorithm for different SCAN pattern. Pareek et al. [5] explained the concept of image encryption using chaotic logistic map and verity of encryption methods are developed for image encryption using this chaotic map. Bibhudendra Acharya et al. [6] proposed a concept to perform image encryption using Hill Cipher. Hill cipher is a type of mono alphabetic poly graphic substitution cipher. Hill cipher is a block cipher that has several advantages such as disguising letter frequencies of the plaintext, its simplicity because of using matrix multiplication and inversion for enciphering and deciphering, its high speed, and high throughput. Decryption requires using the inverse of the matrix K. The inverse matrix K^{-1} of a matrix K is defined by the equation $KK^{-1} = K^{-1} K = I$, where I is the Identity matrix. But the inverse of the matrix does not always exist, and when it does, it satisfies the preceding equation. $K - 1$ is applied to the cipher text, and then the plaintext is recovered. Panduranga [7, 8] explains the basics of encryption using SCAN patterns. In his paper he proposed partial image encryption method using combination of SCAN and mapping images. This method converts a 2D image into a 1D list, and employs a SCAN language to describe the converted result. In this language, there are several SCAN letters. Each SCAN letter represents one kind of scan order. Different kinds of combinations of SCAN letters may generate different kinds of secret images. After determining the combination of SCAN letters, the scheme then generates a SCAN string. This string defines the scan order of the original image.

The rest of this paper is organized as follows: Sect. 2 explains the basics of encryption using SCAN patterns. The concept of phase manipulation technique is explained in Sect. 3. Proposed multiple image encryption method using combination of SCAN and phase manipulation technique is described in Sect. 4. This paper is concluded by providing the summary of the present work in Sect. 5.

2 Image Encryption Using SCAN Patterns

This method converts a 2D image into a 1D list, and employs a SCAN language to describe the converted result. In this language, there are several SCAN letters. Each SCAN letter represents one kind of scan order. Different kinds of combinations of SCAN letters may generate different kinds of secret images. After determining the combination of SCAN letters, the scheme then generates a SCAN string. This string defines the scan order of the original image. Next, this method scans the original image in the determined order and, moreover, encrypts the SCAN string by using commercial cryptosystems. Since the illegal users cannot obtain the correct SCAN string, the original image is therefore secure. There is no image compression in this method. Therefore, the size of the image is very large, and thus it is inefficient to encrypt or decrypt the image directly. Figure 1 shows the basic SCAN patterns, partition patterns and its transformation and resultant encrypted image using different SCAN patterns.

3 Image Modification Using Phase Manipulation Technique

Figure 2 shows the block diagram of phase manipulation technique. In this phase manipulation technique we apply Fourier Transform (FT) to all input images to get phase and magnitude values of each images. Phase values of all images are applied to phase scrambling block to give scrambled phase to each images. This scrambled phase along with the original magnitude is applied to Inverse Fourier Transform (IFT) to obtain modified images. Results of dual image modification using phase manipulation technique is as shown in Fig. 3.

4 Proposed Technique

The concept of proposed technique can be explained by using the block diagram as shown in Fig. 4. Multiple images to be encrypted are applied to phase manipulation block. Fourier Transform is applied to all input images to get phase and magnitude of all images. By using phase scrambling and Inverse Fourier Transform we get Modified images. Finally these modified images are encrypted by using SCAN patterns. Initially dual image encryption is done by using proposed method and finally applied for multiple images by using dual image encryption concept. The results are as shown in Fig. 5.

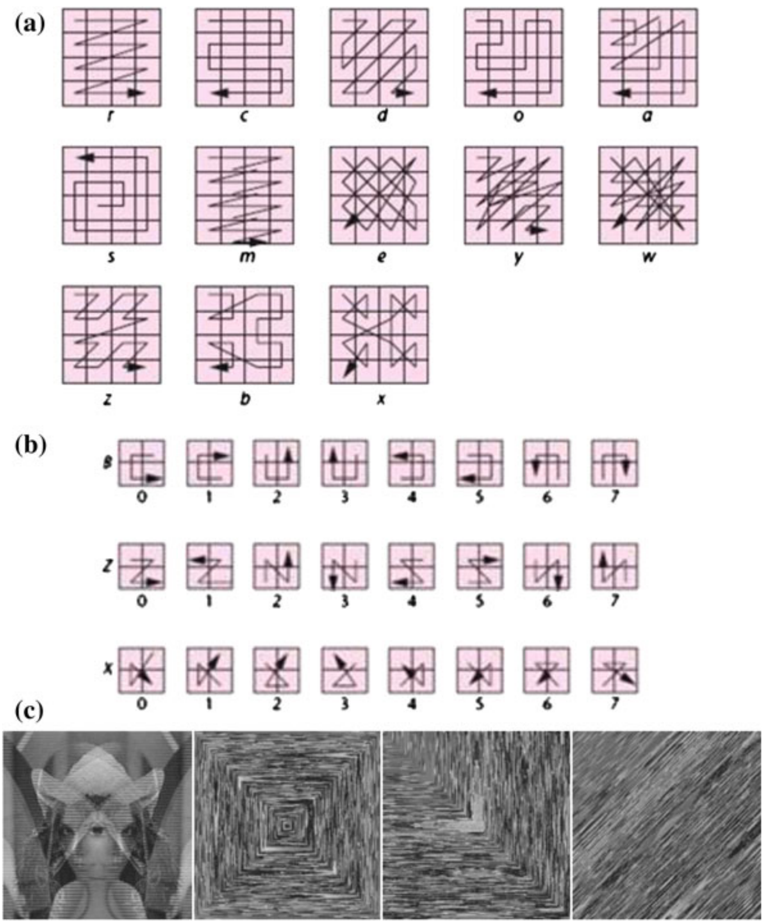


Fig. 1 **a** Basic SCAN patterns, **b** partition patterns and transformations and **c** results using different SCAN

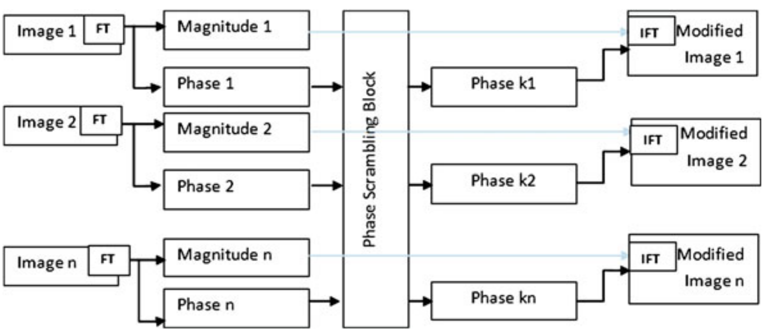


Fig. 2 Block diagram of phase manipulation technique

Fig. 3 **a**, **b** Input images, **c**, **d** modified output images

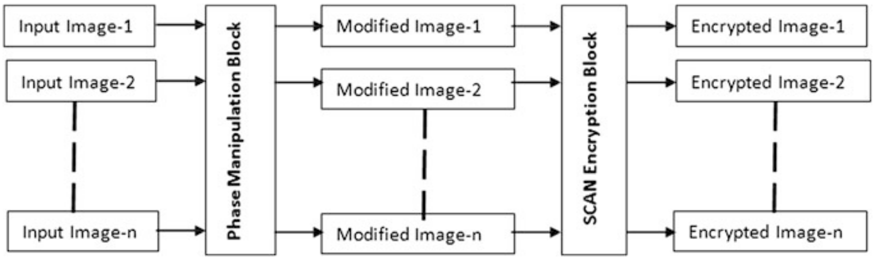


Fig. 4 Block diagram of proposed method

There are different scan patterns to rearrange image pixels to scramble its position like roster scan (r-scan), continues roster scan (c-scan), orthogonal scan (o-scan), diagonal scan (d-scan) and so on. In our proposed method SCAN encryption block consists of d-scan block fallowed by o-scan block. In Fig. 5a, b first two images are original image to be encrypted, second two images are modified image using phase manipulation block and third two images are final encrypted images using SCAN encryption block. In Fig. 5c first image is Lena image second image created by four other images to be encrypted, here both these image can be constructed by multiple images but the size of these resulting two images are same. Second two images are modified image using phase manipulation block and third two images are final encrypted images using SCAN encryption block. Here the resulting images are similar to dual image encryption, but our

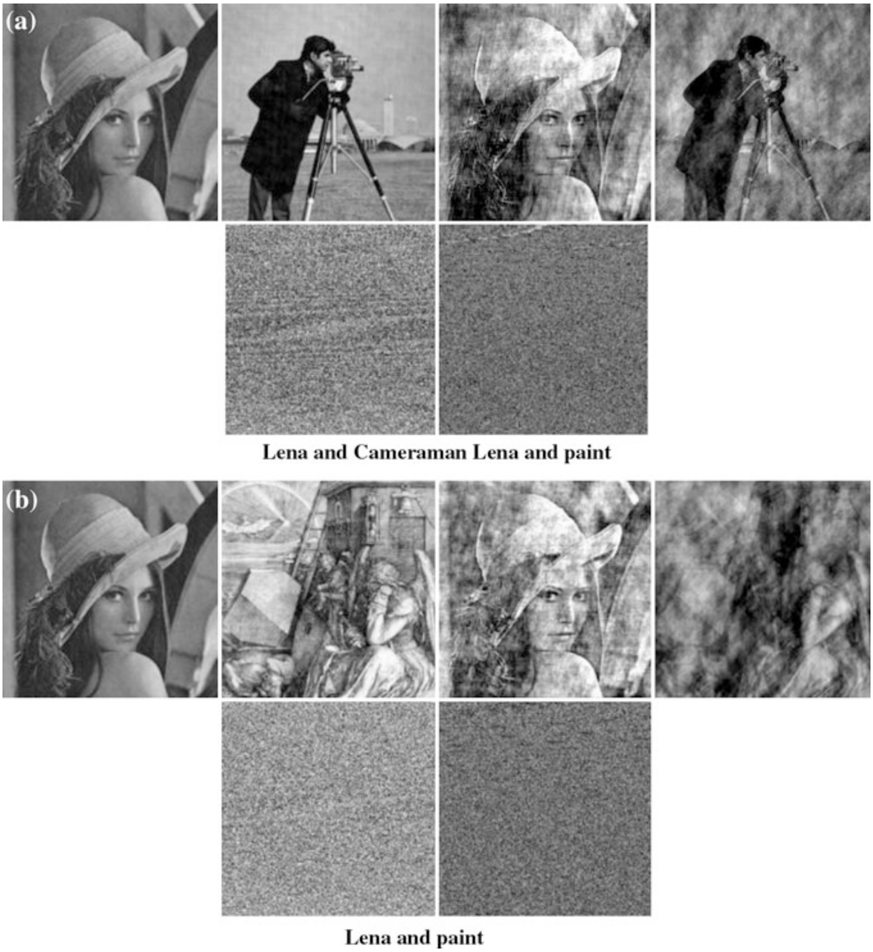


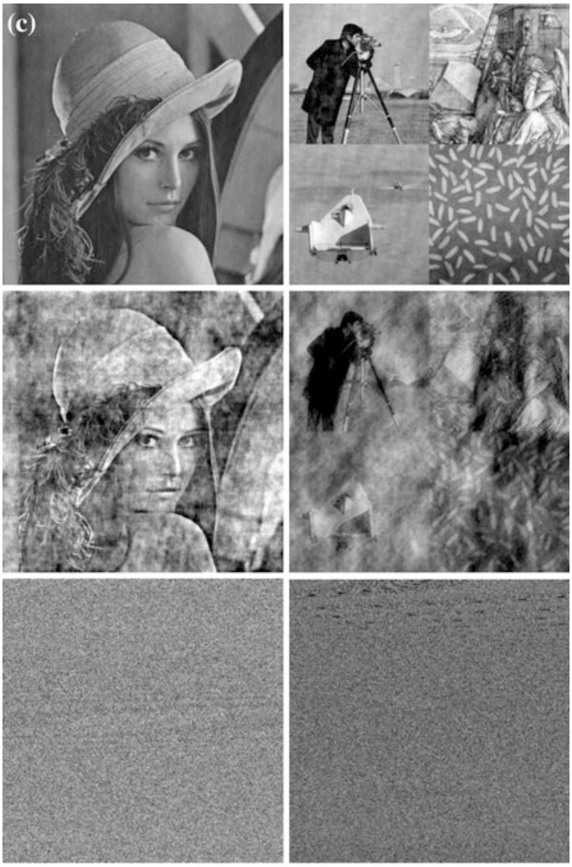
Fig. 5 **a**, **b** Input images, modified images and encrypted images using dual image encryption technique. **c** Input images, modified images and encrypted images using multiple images and dual image encryption technique

proposed technique works for multiple-image encryption where number of input images must be multiple of two.

5 Conclusion

In this paper we are presented a hybrid concept of pixel value manipulation using phase manipulation in frequency domain, and position manipulation in spatial domain using SCAN patterns. From the experiment result we found that the

Fig. 5 continued



Lena and multiple image

combine encryption process gives better results than the individual encryption approaches. This approach can be used for dual image encryption, selective image encryption and multiple image encryptions.

References

1. Chen T-H, Wu C-S (2011) Efficient multi-secret image sharing based on Boolean operations. Sig Process 91:90–97
2. Prasanna SRM et al (2006) An image encryption method with magnitude and phase manipulation using carrier images. IJCS 1(2):132–137
3. Maniccam SS, Bourbakis NG (2004) Image and video encryption using SCAN patterns. Pattern Recogn 37:725–737
4. Kachris C et al (2003) A reconfigurable logic based processor for the SCAN image and video encryption algorithm. IJPP 31(6):489–506

5. Pareek NK et al (2006) Image encryption using chaotic logistic map. Image Vis Comput 24:926–934
6. Acharya B et al (2009) Image encryption using advanced hill cipher algorithm. Int J Recent Trends Eng 1(1):663–667
7. Panduranga HT, Naveen Kumar SK (2011) Hybrid approach to transmit a secrete image. IEEE 978-1-4244-9581-8/11
8. Panduranga HT et al (2010) Hybrid approach for image encryption using SCAN patterns and Carrier Images. IJCSE 2(2):297–300

A Fully Automatic Scheme for Skull Stripping from MRI of Head Scans Using Morphological Neck Breaking Operations

K. Somasundaram and K. Ezhilarasan

Abstract In this paper we propose a fully automatic method for extracting brain portion from T1 weighted MRI of human head scan. The proposed scheme comprise of simple image manipulation methods, intensity thresholding, image binarization, largest connected component analysis, 2D Euclidean distance and morphological operations. Application of our scheme on 20 volumes of MRI data sets shows that the proposed scheme performs better than the existing popular method Brain Extraction Tool (BET) and Brain Surface Extractor (BSE). The proposed scheme gives an average value of 0.944 and 0.970 for the similarity indices Jaccard and Dice.

Keywords Magnetic resonance image · Morphological operations · 2D Euclidean distance · Skull stripping

1 Introduction

Brain is the central organ of our human body. It plays a vital role in the function of human body. It is extremely complex and sophisticated. The brain controls all other organs of the human body through the central nervous system. The human brain consists of three main parts, Cerebrum, Brain stem (medulla) and

K. Somasundaram (✉) · K. Ezhilarasan
Image Processing Lab, Department of Computer Science and Applications,
Gandhigram Rural Institute (Deemed University), Gandhigram 624302,
Dindigul, Tamil Nadu, India
e-mail: ka.somasundaram@gmail.com

K. Ezhilarasan
e-mail: ezhilarasankc@gmail.com

Mohan S. and S. S. Kumar (eds.), *Proceedings of the Fourth International Conference on Signal and Image Processing 2012 (ICSIP 2012)*, Lecture Notes in Electrical Engineering 222,
DOI: 10.1007/978-81-322-1000-9_25,

Cerebellum. The brain is covered by the bones of skull which protects brain from physical shock [1]. Brain abnormalities that occur in the brain tissues are classified into five categories such as degenerative diseases, bleeding, tumors, infections and hormonal disorders. These diseases can be diagnosed by taking images of the tissues in the brain. The images of brain help the neurologist to know the depth of the infections and find suitable interventional methods. The growth of medical imaging is very rapid during recent years in the modern medical field. The medical images are taken using different imaging modalities such as X- rays, CT- scans, Ultra Sound images, Magnetic Resonance Image (MRI) and so on. The MRI is used to study the anatomical structure of soft tissues in human body. This is a non-ionizing, non-destructive and non-invasive method. MRI is taken in three different relaxation times T1-weighted, T2- weighted and Proton Density (PD). MRI is taken in three orientations axial (top to bottom of the head), sagittal (left to right of the head) and coronal (back to front of the head). Each orientation shows the different views of the same organ. The MRI of head scan has many regions which are associated with the brain tissue characteristics. To understand the diseases or problems in brain, physicians need clear perception of brain tissues. Image segmentation aims at splitting an entire image into a set of regions. Segmentation helps to extract uniform areas more compactly. The skull stripping of MRI is a pre-process for several other MRI processing like registration and compression. Manual segmentation of brain from MRI takes more time. An automated segmentation method reduces the segmentation time and avoids operator biases. In recent years many research works were reported in brain segmentation. The research work on brain portion segmentation by an automated method was reported in 1996 [2]. Few popular methods for segmenting brain are Brain Extraction Tool (BET) [3], and Brain Surface Extractor (BSE) [4]. The Brain Extraction Tool (BET) removes the skull from an image, leaving only the region occupied by actual brain tissue. It segments the brain using the dark space between the brain and skull occupied by the cerebrospinal fluid (CSF). Binary image differentiates the darker and brighter regions in the image. The intensity threshold value is computed using histogram. Brain Surface Extractor (BSE) uses edge detecting filters and morphological operations to extract the surface of the brain from the head scans. Few recent algorithms reported for automatic brain segmentation are simplex mesh and histogram analysis [5], Brain Extraction Algorithm [6], 2D region growing [7], 2D-Brain Extraction Method and 3D- Brain Extraction Method [8].

In this paper we propose a new method for segmenting the brain portion from the MRI of human head scans. This method is based on morphological operations [9] and 2D Euclidean distance [10]. The performance of this method is evaluated by estimating the value of Jaccard and Dice coefficients by comparing it with the hand segmented image called gold standard. The remaining part of the paper is organized as follows. The proposed method and materials used are given in Sect. 2, experimental results and discussion are given in Sect. 3 and finally conclusions are given in Sect. 4.

2 Methods and Materials Used

2.1 Materials Used

We used 20 normal volumes of T1 weighted images taken from the Internet Brain Segmentation Repository (IBSR) [11]. Each volume contains 50 to 60 slices and each slice has size of 256 $\times$ 256 pixels. The slices are in coronal view. Table 1 gives details of data sets used for our experiment.

2.2 Proposed Method

The proposed scheme consists of several image processing techniques to segment the brain portion from the head scans of MRI.

First, an intensity threshold value is computed by using Riddler's method and the background is eliminated. Then top-hat filtering is done to diffuse other irrelevant portions from the object. Then a 2D Euclidean distance measure is used to remove narrow connections. Then the largest connected component analysis (LCC) is done to segment the rough brain portion from the MRI of head scans. Finally morphological operations are performed to detect the fine brain portion. The flow chart of the proposed method is shown in Fig. 1.

Table 1 Details of the 20 normal volumes of T1 weighted coronal MR of head scans

Index	Volume Name	Gender	Age at scan taken
1.	1_24	F	35
2.	2_4	F	34
3.	4_8	F	29
4.	5_8	F	20
5.	6_10	M	22
6.	7_8	M	29
7.	8_4	M	27
8.	11_3	M	28
9.	12_3	M	38
10.	13_3	M	32
11.	15_3	M	31
12.	16_3	F	36
13.	17_3	F	29
14.	100_23	M	23
15.	110_3	M	25
16.	111_2	M	27
17.	112_2	M	32
18.	191_3	M	32
19.	202_3	F	28
20.	205_3	F	24

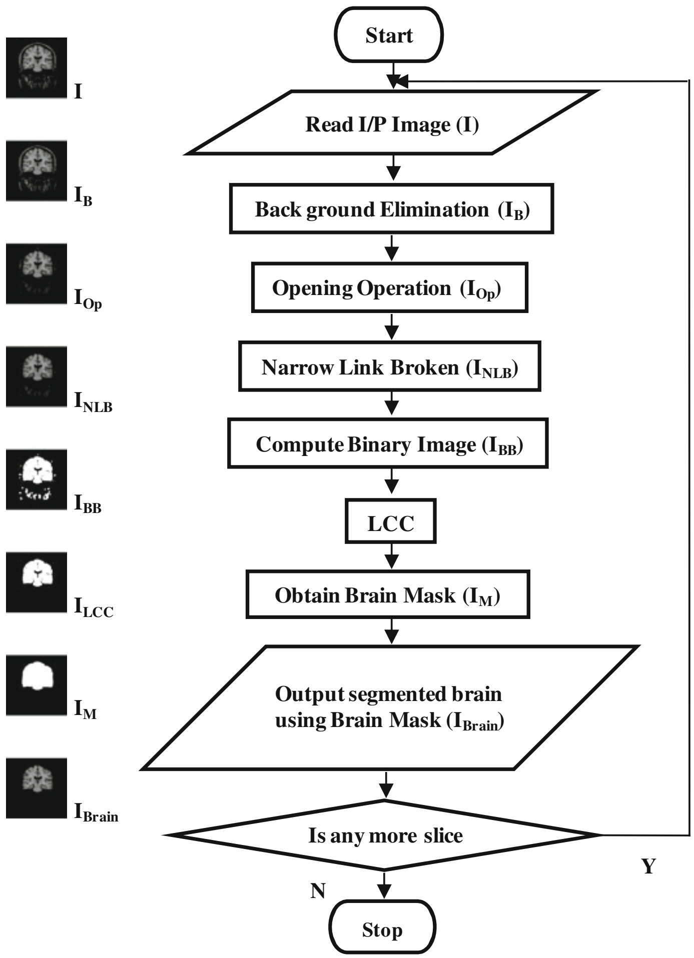


Fig. 1 Flow chart of the proposed scheme

2.2.1 Intensity Threshold

Riddler's method is an iterative one for obtaining intensity threshold value for an image. It generates the optimal intensity threshold value [6]. Intensity threshold helps to remove the background pixels. In this method, an initial threshold value T_1 is computed as:

$$\mathbf{T_1} = \frac{\sum_{i=0}^{N-1} \mathbf{x_i}}{\mathbf{N}} \tag{1}$$

where, x_i is the intensity of the ith pixel in the input image of size m × n and N (= m × n) is the total number of pixels in the given image. Using the threshold value T_1 the pixels in the image are separated into two sets S_1 and S_2 as:

$$\mathbf{f(x_i)} \in \begin{cases} \mathbf{S_1} & \mathbf{if\ x_i > T_1} \\ \mathbf{S_2} & \mathbf{otherwise} \end{cases} \tag{2}$$

Next we compute a new threshold value T_{new} as:

$$\mathbf{T_{new}} = \frac{\frac{\sum_0^{C_1-1} \mathbf{S_1}}{\mathbf{C_1}} + \frac{\sum_0^{C_2-1} \mathbf{S_2}}{\mathbf{C_2}}}{\mathbf{2}} \tag{3}$$

where, C_1 is the number of pixels in S_1 and C_2 is the number of pixels in of S_2.

Equations (2) and (3) are repeated until the values of T_1 and T_{new} are nearer. The final T_{new} is taken as the threshold value T. Using the threshold value T, we generate a background eliminated image I_B:

$$\mathbf{I_B} = \begin{cases} \mathbf{I(x_i)} & \mathbf{if\ I(x_i) \geq T} \\ \mathbf{0} & \mathbf{otherwise} \end{cases} \tag{4}$$

where, I(x) is the input image and i = 0,1,2,……. N-1.

2.2.2 Image Opening

Gray Scale opening (erosion followed by dilation) tends to suppress small bright regions in the image whilst leaving the rest of the image relatively unchanged. The given image I_B is first eroded using gray scale erosion using a structuring element of size 5 × 5, and we get a gray scale eroded image [9] I_{GE} as:

$$\mathbf{I_{GE} = I_B \ominus B} \tag{5}$$

The gray scale erosion operation is performed using the following steps.

i Place the structuring element B over each pixel in the image I_B. (Edges of two pixels are unaltered).

ii For each pixel, find the minimum value of pixels under the mask B and replace the current pixel with the minimum value obtained.

We then perform a gray scale dilation which gives an image I_{Op} as:

$$\mathbf{I_{op} = I_{GE} \oplus B} \tag{6}$$

Figure 2 shows the process of gray scale erosion by 5 × 5 size mask. The gray scale dilation is shown in Fig. 3.

The gray scale dilation is done using the following steps.

i Successively place the structuring element B over each pixel in the image I_{GE}.

ii For each location, find the maximum value of pixel under the mask of the 5 × 5 structuring element and replace the current pixel value by the maximum value.

Fig. 2 Gray scale Erosion

197	168	127	186	152
168	138	168	198	168
148	157	168	125	168
128	138	168	157	168
125	138	129	168	148

Fig. 3 Gray scale Dilation

197	168	125	186	152
168	138	168	190	168
148	157	168	125	168
128	138	168	157	168
125	138	129	168	148

This process is repeated until all the pixels in the image are processed by the opening operation. The resulting image after dilation is taken as I_{Op}.

2.2.3 Breaking Narrow Links

In T_1 weighted images some non brain tissues have similar intensities which lies in-between gray matter and cerebrospinal fluid (CSF). Such narrow connections are removed using 2D Euclidean distance measure. The image I_{Op} may have such a narrow connection, and are removed by using the following procedure.

Initially, we define a window of size 33 × 33, centered at the object pixel x ($I_{Op}(x) > 0$) as seed and move the mask over the entire image I_{Op} row by row. Starting from the current pixel, a test is made whether any 5 consecutive background pixels appear along the 16 pixel distance. If such a pattern occurs, then the Euclidean distance d_i is computed as:

$$\mathbf{d_i}=\sqrt{(\mathbf{x_1}-\mathbf{x_2})^2+(\mathbf{y_1}-\mathbf{y_2})^2} \tag{7}$$

where (x_1, y_1) and (x_2, y_2) are the co-ordinates of the object pixel and the first occurrence of background pixel in that direction and i = 1,…8 (left, right, top, bottom, left top diagonal, right bottom diagonal, right top diagonal and left bottom diagonal) are the direction indices as shown in Fig. 4. If no such pattern occurs,

Fig. 4 Breaking narrow links

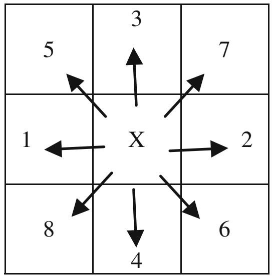
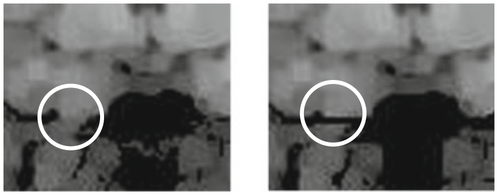


Fig. 5 Shows Narrow link breaking. **a** Initial image with neck (rounded in white), **b** Narrow neck link broken from (**a**) (appears in block)



(a) **(b)**

then d_i is set as 0. This is repeated for all the 8 direction for each object pixel and the image I_{NLB} is computed as:

$$\mathbf{I_{NLB}(x)} = \begin{cases} \mathbf{0}, & \mathbf{if} \begin{pmatrix} (\mathbf{d_1} \neq \mathbf{0\&\&d_2} \neq \mathbf{0})\mathbf{P}(\mathbf{d_3} \neq \mathbf{0\&\&d_4} \neq \mathbf{0}) \\ (\mathbf{d_5} \neq \mathbf{0\&\&d_6} \neq \mathbf{0})\mathbf{P}(\mathbf{d_7} \neq \mathbf{0\&\&d_8} \neq \mathbf{0}) \end{pmatrix} \\ \mathbf{I_{Op}(x)}, & \mathbf{otherwise} \end{cases} \tag{8}$$

By operating Eq. (8), the weakly connected necks are broken. The Fig. 5a shows the neck and Fig. 5b shows the "broken neck".

2.2.4 Amplify Threshold Value

The image (INLB) includes all brain tissues as well as some non brain tissues with the same intensity. The amplified threshold TA for the image INLB is computed as in [12].The constant value 0.9 is selected after testing with various values which ranges between 0.1 and 0.9, and found that 0.9 gives good result for this process.

$$\mathbf{T_A} = \mathbf{T} + \mathbf{0.9}(\mathbf{T} - \mathbf{S}) \tag{9}$$

where,

$$\mathbf{S} = \frac{\sum_{\mathbf{i=0}}^{\mathbf{N-1}} \mathbf{I_{NLB}(x_i)}}{\mathbf{N}} \tag{10}$$

where N is the total number of pixel in the image I_{NLB}. Using T_A, we generate binary image I_{BB} as:

$$\mathbf{I_{BB}} = \begin{cases} \mathbf{1} & \mathbf{if\ I_{NLB}(x_i) \geq T_A} \\ \mathbf{0} & \mathbf{otherwise} \end{cases} \tag{11}$$

2.2.5 Largest Connected Component

It is known that the brain is the largest portion in the middle slice of head region. Therefore we employ Largest Connected Component (LCC) algorithm [6] to extract the brain from the head scan. The binary brain image I_{LCC} is obtained after applying LCC on I_{BB}.

2.2.6 Morphological Dilation

Dilation process is one of the primitive morphological operations which enlarges the area surrounded by the object. We perform this dilation on binary image to recover the pixels lost during the amplification of threshold value. For that we use structuring element E of size 3×3, 5×5 and 7×7. We found from several trail experiments, that the structuring element of size 5×5 gives the best results. Hence we have set the structuring element to 5×5 pixel. The dilation is performed on the image I_{LCC} to get the brain mask (I_M).

$$\mathbf{I_M = I_{LCC} \oplus E} \tag{12}$$

The brain mask I_M is used to extract the brain portion I_{Brian} from the input image I.

3 Results and Discussion

We carried out experiments by applying the proposed scheme on the 20 volumes of data sets described in Table 1. For illustration, an original volume of MR images and the brain portion extracted from them are shown in Fig. 6.

We also make quantitative performance analysis for the proposed scheme by computing the similarity measures, Jaccard index (J) [13] and Dice (D) [14] coefficient. The hand segmented "gold standard" images available in IBSR are used for comparing J and D. The Jaccard index is given by:

$$\mathbf{J} = \frac{|\mathbf{A} \cap \mathbf{B}|}{\mathbf{A} \cup \mathbf{B}} \tag{13}$$

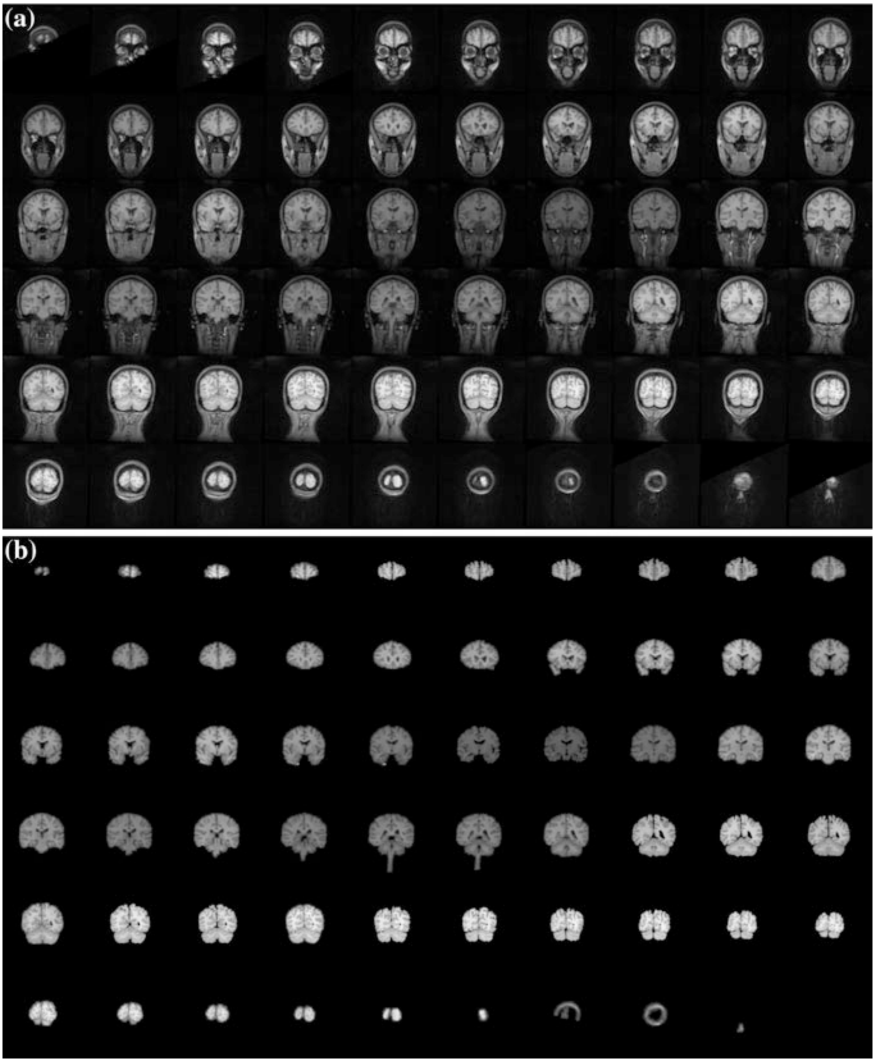


Fig. 6 Brain portion extracted using the proposed scheme, **a** Original image (Volume 205_3). **b** Extracted brain portion from (**a**)

and Dice coefficient D is given by:

$$\mathbf{D} = \frac{\mathbf{2}|\mathbf{A} \cap \mathbf{B}|}{|\mathbf{A}| + |\mathbf{B}|} \tag{14}$$

Table 2 Computed values of J, D, FPR and FNR using our scheme on 20 data sets

Index	Proposed Method				BSE				BET			
	J	**D**	**FPR**	**FNR**	**J**	**D**	**FPR**	**FNR**	**J**	**D**	**FPR**	**FNR**
1	0.946	0.972	0.022	0.958	0.9	0.947	0.048	0.052	0.758	0.862	0.241	0.0005
2	0.958	0.978	0.013	0.008	0.92	0.959	0.059	0.02	0.808	0.894	0.169	0.023
3	0.936	0.967	0.022	0.01	0.905	0.95	0.055	0.04	0.828	0.906	0.151	0.021
4	0.951	0.975	0.018	0.012	0.712	0.832	0.044	0.244	0.772	0.871	0.204	0.025
5	0.962	0.981	0.018	0.001	0.718	0.836	0.259	0.023	0.667	0.8	0.317	0.017
6	0.956	0.977	0.015	0.007	0.891	0.942	0.046	0.062	0.683	0.812	0.315	0.002
7	0.96	0.98	0.016	0.004	0.914	0.955	0.046	0.04	0.778	0.875	0.218	0.004
8	0.945	0.971	0.027	.0005	0.914	0.955	0.075	0.011	0.853	0.921	0.147	0.001
9	0.908	0.937	0.037	0.029	0.87	0.931	0.117	0.013	0.788	0.881	0.212	0.001
10	0.919	0.944	0.048	0.014	0.889	0.941	0.095	0.016	0.857	0.923	0.143	0.0003
11	0.896	0.939	0.027	0.029	0.909	0.952	0.039	0.052	0.684	0.812	0.297	0.019
12	0.948	0.972	0.023	0.003	0.894	0.944	0.039	0.067	0.54	0.701	0.457	0.002
13	0.936	0.966	0.029	0.003	0.908	0.952	0.076	0.016	0.578	0.732	0.422	0.0003
14	0.953	0.975	0.013	0.011	0.918	0.957	0.066	0.017	0.827	0.905	0.167	0.006
15	0.945	0.971	0.025	0.003	0.913	0.954	0.079	0.008	0.777	0.875	0.222	0.001
16	0.949	0.986	0.009	0.005	0.915	0.956	0.071	0.014	0.855	0.922	0.143	0.002
17	0.952	0.976	0.023	0.001	0.906	0.95	0.084	0.01	0.779	0.876	0.221	0.0001
18	0.962	0.98	0.017	0.002	0.927	0.962	0.055	0.017	0.846	0.916	0.154	0.0003
19	0.955	0.976	0.019	0.004	0.919	0.958	0.071	0.01	0.858	0.924	0.14	0.001
20	0.959	0.979	0.02	0.001	0.926	0.961	0.062	0.012	0.711	0.831	0.089	0.2
Avg.	**0.945**	**0.97**	**0.022**	**0.007**	**0.888**	**0.94**	**0.074**	**0.037**	**0.762**	**0.862**	**0.221**	**0.016**

where, A is the pixels in standard image and B is the pixels in the segmented image by our scheme. We also compute the parameters False Positive Rates (FPR) and False Negative Rates (FNP) [15] using:

$$\mathbf{FPR} = \frac{|\mathbf{A} \cap \mathbf{B^C}|}{\mathbf{A} \cup \mathbf{B}} \tag{15}$$

$$\mathbf{FNR} = \frac{|\mathbf{A^C} \cap \mathbf{B}|}{\mathbf{A} \cup \mathbf{B}}$$

From the Table 2 we note that the proposed method gives best values 0.945 and 0.970 for Jaccard and Dice similarity, and the lowest values 0.022 and 0.007 for FPR and FNR respectively, compared to BSE and BET.

4 Conclusions

In this paper we have proposed a new fully automatic scheme for extracting brain portion from T1 weighted MRI of human head scans. The proposed method does not require any preprocessing like noise removal. Experimental results by applying the proposed method on 20 volumes of MRI obtained from the publically available repository IBSR show that the proposed method performs better than the existing popular methods BET and BSE. The proposed scheme gives an average value of 0.944 for Jaccard coefficient and 0.970 for Dice coefficient.

Acknowledgments This work is funded by the University Grants Commission, New Delhi, through the Grant No: F No 37/154/2009(SR). The authors would like to thank the Internet Brain Segmentation Repository (IBSR) for providing 20 volumes of MRI brain images.

References

1. Kalaiselvi T (2011) Brain portion extraction and brain abnormality detection from magnetic resonance imaging of human head scans. Pallavi publications India pvt.ltd, Erode. ISBN 978-93-80406-76-3
2. Kapur T, Grimson WEL, Wells WM 3rd, Kikinis R (1996) Segmentation of brain tissue from magnetic resonance images. Med Image Anal 12:109–127
3. Smith S (2002) Fast robust automated brain extraction. Hum Brain Mapp 17:143–155
4. Sandor S, Leahy R (1997) Surface-based labeling of cortical anatomy using a deformable atlas. IEEE Tans on Med Imaging 16:41–54
5. Galdames FJ, Jaillet F, Perez CA (2012) An accurate skull stripping method based on simplex meshes and histogram analysis in magnetic resonance images. J Neurosci Methods 206:103–119
6. Somasundaram K, Kalaiselvi T (2011) Automatic brain extraction methods for T1 magnetic resonance images using region labeling and morphological operations. Comput Biol Med 41:716–725
7. Park JG, Lee C (2009) Skull stripping based on region growing for magnetic resonance brain images. NeuroImage 47:1394–1407
8. Somasundaram K, Kalaiselvi T (2010) Fully automatic brain extraction algorithm for axial T2-weighted magnetic resonance images. Comput Biol Med 40:811–822
9. Gonzalez RC, Woods RE (2009) Digital image processing, 3rd edn. Dorling Kindersley (India) Pvt. Ltd., New Delhi, ISBN 978-81-317-2695-2
10. Saito T, Toriwaki J-I (1994) New algorithms for euclidean distance transformation of an n-dimensional digitized picture with applications. Pattern Recognit 27:1551–1565
11. http://www.cma.mgh.harvard.edu/ibsr/
12. Shan Z, GH Y, Liu J (2002) Automated histogram-based brain segmentation in T1-weighted three-dimensional magnetic resonance head images. Neuroimage 17:1587–1598
13. Jaccard P (1912) The distribution of flora in the alpine zone. New Phytol 11:37–50
14. Dice L (1945) Measures of the amount of ecologic association between species. Ecology 26:297–302
15. John C, Kevin W, Emma L, Chao C, Barbara P, Declan J (2007) Statistical morphological skull stripping of adult and infant MRI data. Comput Biol Med 37:342–357

A Method for Segmentation Radiographic Images with Case Study on Welding Defects

Alireza AzariMoghaddam and Lalitha Rangarajan

Abstract Segmentation is one of the most difficult tasks in image processing, particularly in the case of noisy or low contrast images such as radiographic images of welds. In the present study we have segmented defects in radiographic images of weld. The method applied for detecting and discriminating discontinuities in the radiographic weld images. Two Dimensional Left Median Filter (2D-LMF) has been used for enhancing the images. We compared the performance of this method with Mean Shift. Results exhibited the applied method was more effective than Mean Shift in noisy and low contrasted radiographic images of weld.

Keywords Segmentation · Radiographic Image · Weld defect · Mean shift

1 Introduction

Welding is process of joining which is used in larger applications of mechanical engineering [1]. The inspection of welds is a very important task for assuring safety and reliability in industries. Non-Destructive Testing (NDT) techniques have been employed to test a material for surface or internal flaws, without interfering in any way with its suitability for service. Radiography seems to be the most effective method and the experts are able to identify most types of defects in

A. AzariMoghaddam (✉) · L. Rangarajan
Department of Study in Computer Science, University of Mysore, Mysore, India
e-mail: arazarim@gmail.com

L. Rangarajan
e-mail: lali85arun@yahoo.co.in

Mohan S. and S. S. Kumar (eds.), *Proceedings of the Fourth International Conference on Signal and Image Processing 2012 (ICSIP 2012)*, Lecture Notes in Electrical Engineering 222,
DOI: 10.1007/978-81-322-1000-9_26,

the images produced by this method. The method is based on the fact that the defective areas absorb more energy and thus the defects appear darker in the image [2].

The progress in computer science and the artificial intelligence techniques have allowed the defect detection and classification to be carried out by using digital image processing and pattern recognition tools [3]. Computer vision applications often require segmentation of digital imagery into semantically meaningful regions. The segmented regions can provide a basis for subsequent tasks such as object detection and recognition [4].

The segmentation of defects detection has been addressed by several researches. In most of the papers, an automated vision system has been introduced to detect the welding defects from the radiographic images, using various image processing techniques. Necessary information such as length, width, area and perimeter of the defects are calculated [5–14].

Research in this area continues largely because no satisfactory results that allow the detection of the totality of defects without false alarms are available. Moreover, it is not possible to determine which lines of investigation will improve the overall result, because each of them is needed to be improved.

In this article we will focus on welding defects using segmentation techniques applied to the radiographic images.

We enhance radiographic images by Two Dimensional Left Median Filter (2D-LMF) [15, 16] and propose a method for segmentation of weld defects. We compare the performance of this method with Mean Shift.

The mean shift estimates of the gradient of a density function and the associated iterative procedure of mode seeking have been developed by Fukunaga and Hostetler in [17]. It is employed in the joint, spatial-range (value) domain of gray level and color images for discontinuity preserving filtering and image segmentation.

The paper is organized as follows: 2D-LMF is described in Sect. 2, the proposed algorithm is shown in Sect. 3, experimental results exhibited in Sect. 4. Section 5 concludes the paper.

2 Two Dimensional Left Median Filter

The radiographic images are contaminated with noise and are also blurred. In order to improve the image for observation and accurate analysis, various digital image processing techniques can be applied. Noise removal is required for improving the quality of the image in order to better recognize the defects. Different noise removal filters are used [18]. In digital image processing, filtering is the most common and basic operation, because the results of filtering directly influence all the following operations such as edge detection, image enhancement, etc. [19]. We use Two Dimensional Left Median Filter for this purpose.

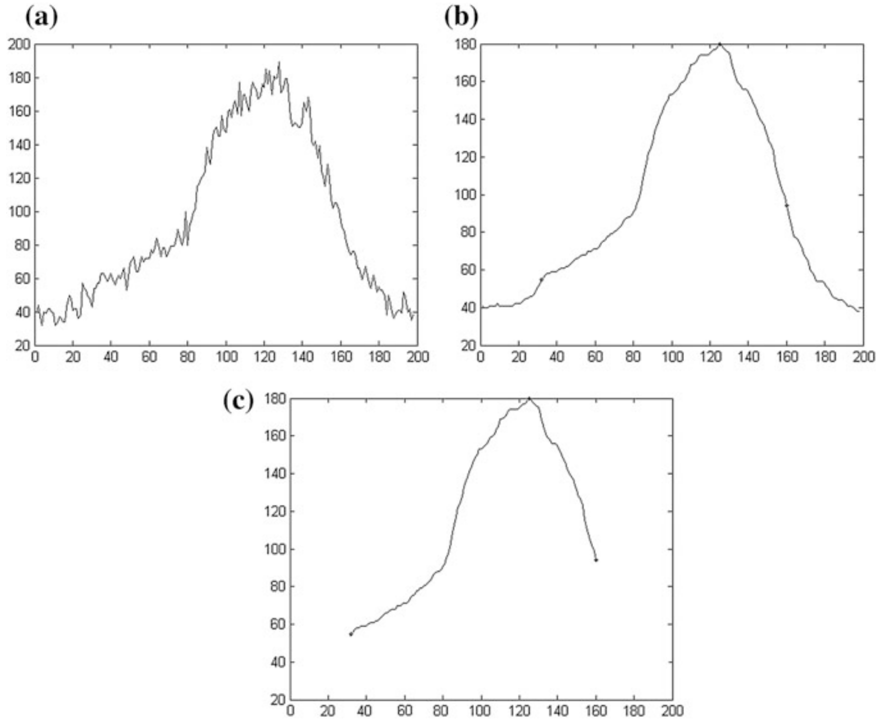


Fig. 1 Sample line profile without defect. **a** Original profile. **b** After 2D-LMF. **c** Area of interest in the profile

In the 2D-LMF, a square window of size p X p (p is an odd number) is considered. The centre pixel in the scan window is to be de-noised. The value of the central pixel in the scan window is replaced by the mean of q pixels of the scan window.

2.1 Methodology

Algorithm 2D- Left Median Filter Algorithm
Input I = A radiographic image (m X n),
p = size of 2D – window
Output 2D-Left Median Filter(2DLMF) image of size (m X n)

Algorithm
Step 1 Compute q = (p + 1)/2
Step 2 Create a matrix J of size ((m + p-1) X (n + p-1)) from I.
Step 3 Consider scan window W of size p X p for elements ($J_{i,j}$),where $J_{i,j}$ is center of scan window, i ranges from q to m-q and j ranges q to n-q.

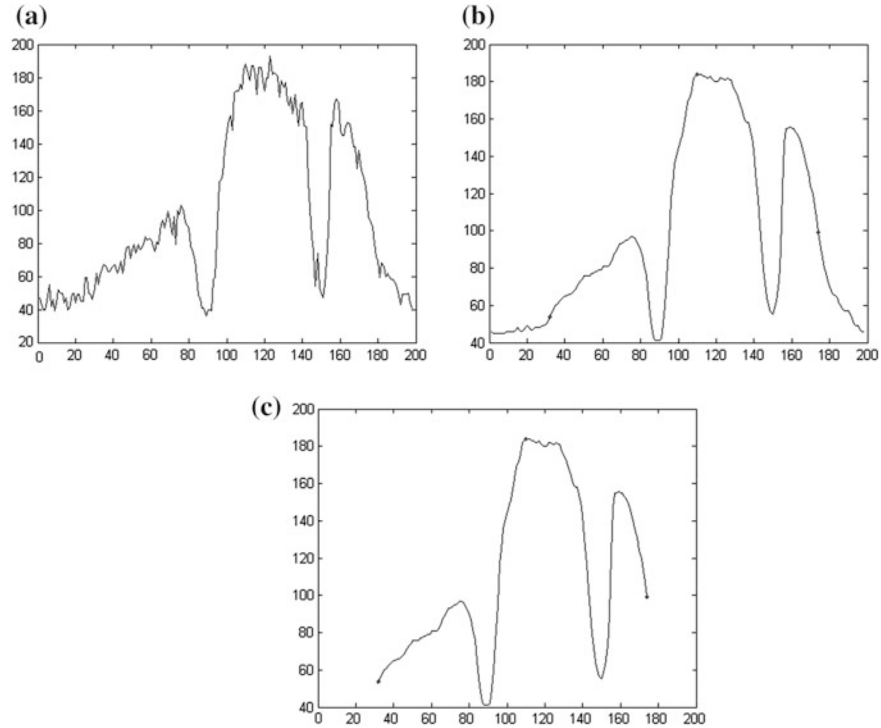


Fig. 2 Sample line profile with defect. **a** Original profile. **b** After 2D-LMF. **c** Area of interest in the profile

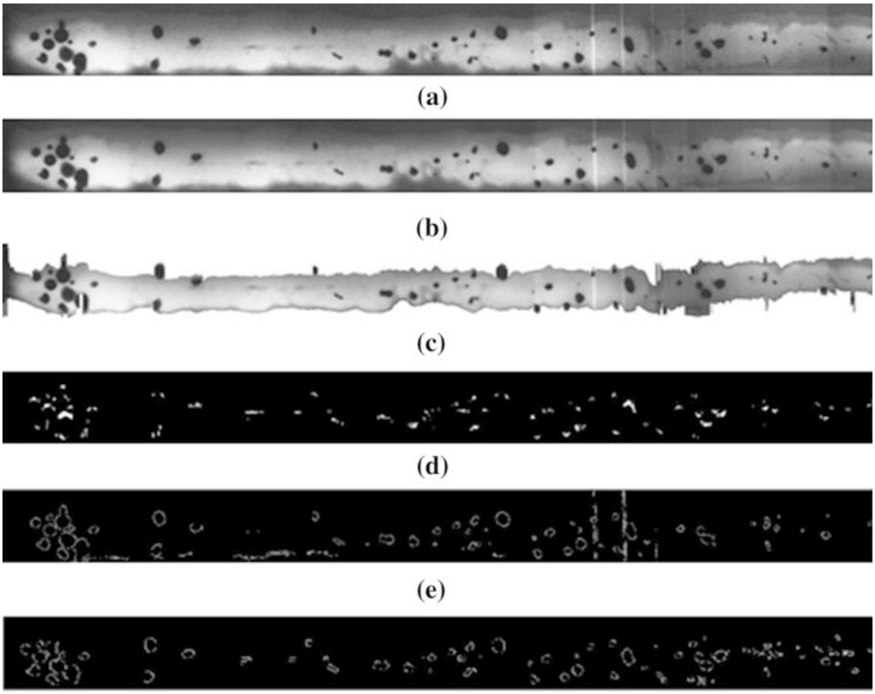
Step 4 Sort elements of W : $\left(X_1, \ldots, X_{med-1}, X_{med}, X_{med+1}, \ldots, X_{pxp}\right)$

Step 5 Define $2DLMF_{i,j} = \left(\sum_{(med-q+1)}^{med} X_k \right) / q$

3 Our Method

From the observation of the horizontal line profiles of weld images without defects, we found one common feature of almost all the line profiles: each profile has a bell shape like a Gaussian curve, as shown in Fig. (1). If there are any defects in a profile, the curve shape has changed, as shown in Fig. (2).

In this research, we have improved our previous algorithm [20] that checks images line by line for finding the defects. We have separated the bell shape to two sections, increasing gray values (IG) and decreasing gray values (DG) sections. As a first step, we found Region of Interest by using mean and variance of elements of



(a)

(b)

(c)

(d)

(e)

(f)

Fig. 3 Sample image with defects. **a** Original. **b** After filtering by 2D-LMF. **c** Area of Interest. **d** Method in [20]. **e** Mean shift. **f** Proposed method

each line. In IG section, any decreasing gray value is an anomaly, and in DG section, any increasing gray value is an anomaly.

Then we detect these anomalies by using threshold in the second step. Approved size of defects based on American Petroleum Institute Standard (API 1104) has been defined, thus: Maximum diameter of Porosity shall not exceed 2.4 mm and also the width of an Elongated Slag Inclusion indication shall not exceed 1.6 mm, and bigger sizes will be considered as a defect.

There are 2.8×2.8 pixels in each square millimeter of our radiographic images. We tested our algorithm with window sizes 3×3 and 2×4 for finding defects. We found window size 2×4 as the optimal size for detection of defect.

4 Experimental and Result

We tested the proposed approach with gray-scale radiographic images; with gray values 0–255. We tested the algorithm with 32 radiographic images, which include 73,284 horizontal lines. In the selected images 20 % have no defects and 80 % are defective. In preprocessing stage, we used two dimensional left median filters (2D-LMF). We also consider a square window of size 9×9 for filtering. With our method, 98 % of images with defects have been segmented correctly, but for images without defects, the error rate is 15 %. We have compared the performance of the proposed method with [20] which had 27 % error and Mean Shift Fig. 3.

The experimental result shows that the proposed method is effective and feasible to segment and locate defects in noisy and low contrasted radiographic images of weld.

5 Conclusion

Our study has improved the previous segmentation method for defects detection in radiographic images of weld. We used two dimensional left median filter for de-noising the radiographic images. We compared the performance of this segmentation method with [20] and Mean Shift. Experimental results show that, our segmentation methodology has a high successful detection in radiographic images with defects. The proposed method can reduce the working effort of human being and increase the defect detection efficiency. We have to classify weld defects in radiographic images in future studies.

References

1. Vijay RR, Anand RS (2010) Comparison of different segmentation techniques in detection of flaws in weldments. IEEE 2:673–677
2. Ioannis V, Dimitrios K (2010) Multiclass defect detection and classification in weld radiographic images using geometric and texture features. Expert Syst Appl 37(12):7606–7614
3. Faiza M, Naraa N, Archa BG (2007) Unsupervised algorithm for radiographic image segmentation based on the gaussian mixture model. The international conference on computer as a tool, Warsaw, pp 289–293
4. Qin AA, David AC (2010) Multivariate image segmentation using semantic region growing with adaptive edge penalty. IEEE Trans Image Proc 19(8):2157–2170
5. Damin L, Warren TL (1996) Applications of fuzzy K-NN in weld recognition and tool failure monitoring. IEEE 222–226
6. Gang W, Liao WT (2002) Automatic identification of different types of welding defects in radiographic images. NDT and E International 35:519–528
7. Miguel C, Domingo M (2004) Segmentation of welding defects using a robust algorithm. Mater Eval 62:1142–1147
8. Nacereddine N, Zelmat M, Belarfa SS, Tridi M (2005) Weld defect detection in industrial radiography based digital image processing world Academy of science, engineering and technology, PWASET 2:145–148
9. Gao W, Tang N, Mu X (2007) A novel algorithm for detecting air holes in steel pipe welding based on hopfield neural network, eighth acis international conference on software engineering, artificial intelligence, networking, and parallel/distributed computing, Las Vegas, pp 79–83

10. Amin ES (2008) Application of artificial neural networks to evaluate weld defects of nuclear components. J Nucl Radiat Phys 3(2):83–92
11. Abdelhak M, Fakhita R (2009) Fast segmentation method for defects detection in radiographic images of welds. IEEE 857–860
12. Juna Z, Rafael V, Ramon R (2010) An adaptive-network-based fuzzy inference system for classification of welding defects. NDT E Int 43(3):191–199
13. Romeu RS, Marcio HSS, Ivan CS, Antonio AC, Joao MAR (2002) Contribution to the development of a radiographic inspection automated system. J Nondestr Test 7(12):1–8
14. Elineudo PM, Romeu RS, Antonio AC, Marcio HSS, Joao MAR (2003) Welding defecets pattern recognition. In: TOFD signals using linear classifier implemented by neural networks, The 3th Pan-American conference for nondestructive testing Rio
15. Alireza AM, Lalitha R (2011) Enhancing radiographic images using two dimensional left median filter, Third national conference on computer vision, pattern recognition, image processing and graphics (NCVPRIPG2011) Hubli, pp 204–206
16. Alireza AM, Lalitha R (2012) A novel algorithm for de-noising radiographic images. Int J Image Graph Signal Proc 4(6):22–28
17. Keinosuke F, Larry DH (1975) The estimation of the gradient of a density function, with applications in pattern recognition. IEEE Trans Inf Theory 21(1):32–40
18. Amod PR, Ravindran VR (2009) Comparison of different ANN techniques for automatic defect detection in X-ray images, international conference of emerging trends in electronic and photonic devices and systems(ELECTRO) Varanasi, pp 193–197
19. Ashraf A, Hassan M, Ashour M, El Mashade M (2009) Image denoising based on spatial filters an analytical study. IEEE 245–250
20. Alireza AM, Lalitha R (2011) A method for detection welding defects in radiographic images. Int J Mach Intell, 3:307–309

Feature Extraction Using DWT with Application to Offline Signature Identification

Suvarna Joshi and Abhay Kumar

Abstract Handwritten signature is most widely accepted biometrics for person identification. This paper proposes a novel algorithm for offline handwritten signature recognition. Target of this research is to present signature recognition based on coded wavelet coefficient. It works at global level for extraction of discriminate signature features using wavelet transform. Before extracting the features, preprocessing of a scanned handwritten signature image is necessary to isolate the signature part and to remove any unwanted background present. Wavelet transform has been used to extract features from preprocessed signature images. Wavelet coefficients are extracted from detail part of handwritten signature and further wavelet coefficients are coded. Wavelet coefficient coding results in image compression. This causes reduced feature vector size. Hamming distance has been used to find out distance between test signature pattern and training signature pattern. Experiments are carried on signature database for 56 users each of 24 genuine and 9 skilled forgery signatures. One more experiment is carried out on gathered database. Recognition success rate for genuine signatures is 95 %. FAR of proposed algorithm is about 0.22.

Keywords Offline signature identification · Wavelet transform · Detail coefficient

S. Joshi (✉)
Department of Electronics and Tele, UCOER, Pune, Maharashtra, India
e-mail: suvarnaj2@email.com

A. Kumar
Department of Electronics, Devi Ahilya University, Indore, Madhyapradesh, India
e-mail: dr.abhaykumar@gmail.com

Mohan S. and S. S. Kumar (eds.), *Proceedings of the Fourth International Conference on Signal and Image Processing 2012 (ICSIP 2012)*, Lecture Notes in Electrical Engineering 222,
DOI: 10.1007/978-81-322-1000-9_27,

1 Introduction

Biometric recognition or simply, biometrics can define as science of automatic recognition of an individual based on his/her physiological and/or behavioral characteristics. By using biometrics, it is possible establish an individual's identity based on, ''who he/she is,'' rather than by, ''what he/she possesses'' (e.g., an ID card) or ''what he/she remembers'' (e.g., a password) [1]. The security requirements of today's society have placed biometrics at the centre of an ongoing debate concerning its key role in a multitude of applications [2]. Biometrics measures unique characteristics of human to recognize person identity. Common physiological traits include face, iris, fingerprint, hand or palm geometry, retina etc., while behavioral characteristics include signature, voice (which also has a physical component), keystroke pattern, and gait. Banks and Government bodies recognize signatures as a legal means of authentication. Handwritten Signature is one of most important widely accepted human attribute for authorizing person. There is growing demand for fast and accurate human identification. A comparison of signature verification with other recognition technologies (fingerprint, face, voice, retina, and iris scanning) reveals that signature verification has several advantages as an identity verification mechanism. Firstly, signature analysis can only be applied when the person is conscious and willing to write in the usual manner, although it is possible that individuals may be forced to submit the handwriting sample. To give a counter example, a fingerprint may also be used when the person is in an unconscious (e.g., drugged) state. Forging a signature is deemed to be more difficult than forging a fingerprint, given the availability of sophisticated analyses [3]. Unfortunately, signature verification is a difficult discrimination problem since a hand-written signature is the result of a complex process depending on the physical and psychological conditions of the signer, as well as the conditions of the signing process [4]. Signature is a strong variable entity and its verification, even for human experts, is not a trivial matter. Signature plays very important role in human identification in different areas such as banks, financial transaction etc., Lot of researchers have carried out study on handwritten signature identification but still it is very much challenging problem due to large intra class variations. We have presented novel and efficient methodology for signature recognition using DWT based coded feature vectors. Proposed methodology has reported good results for FAR as 0.22. This paper is organized as follows: Sect. 2 presents the work carried by several researchers. Section 3 provides proposed system in detail. Section 4 describes feature extraction algorithm. Section 5 presents details about the classifier. Section 6 reports the experimental results. The paper ends with concluding remarks.

2 Literature Survey

Signature verification can be carried out by using two method online verification and offline verification. Online verification deals with sequential data, such as handwriting speed and pen pressure, pen movement angle etc., obtained with a special device called digital tablet. Offline verification system uses features obtained from scanned handwritten signature images. Offline systems are useful in automatic verification of signatures found on bank checks and documents. Automatic online handwritten signature verification system is presented for authentication of signatures on Australian passports. In this system, fuzzy modeling has been used for developing signature recognition [5]. Deng et al. [6] has proposed a novel approach to off–line handwritten signature verification. Proposed method includes application of a closed–contour tracing technique which uses wavelet–based feature extractor to extract complete and stable features from multiresolutional signals Ghandali and Moghaddam [7] have proposed an language dependent off-line Persians signature identification and verification using Image registration, DWT (Discrete Wavelet Transform) and fusion. They used DWT for features extraction and Euclidean distance for comparing features. Larkins and Mayo have presented off-line signature verification method that is based on Adaptive Feature Threshold (AFT) [8]. They have used combination of spatial pyramid and equimass sampling grids to improve representation of a signature based on gradient direction. In classification phase, they used DWT and graph matching methods. Fakhlai and Pourreza [9] proposed an offline signature recognition approach based on three different kinds of feature extractors—wavelet, curvelet and contourlet transform. The curvature and orientation of a signature image was used as feature. They utilized Support Vector Machine (SVM) as a tool to evaluate the performance of the proposed methods. Prakash and Guru[10] proposed an approach for offline signature verification based on score level fusion of distance and orientation features of centroids. The proposed method used symbolic representation of offline signatures using bi-interval valued feature vector. Distance and orientation features of centroids of offline signatures were used to form bi-interval valued symbolic feature vector for representing signatures. Vargas et al. [11] proposed an offline signature verification system based on grey level information using texture features. Proposed algorithm processes the co-occurrence matrix and local binary pattern for feature extraction. Genuine samples and random forgeries were used to train an SVM model. Random and skilled forgeries were used for testing. For skilled forgeries, they were able to achieve an EER of 12.82 %.

3 Proposed Methodology

Proposed methodology consists of three important steps image preprocessing, feature extraction, feature classification. Block Diagram of offline recognition system algorithm is as shown in Fig. 1. After preprocessing we have extracted significant multiresolution features using DWT.

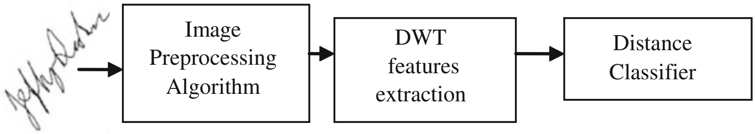
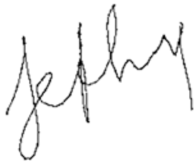


Fig. 1 Offline recognition system algorithm

Fig. 2 Cropped signature image



Handwritten signature images are first preprocessed to remove unwanted background. Signature preprocessing step are as follows

Step 1 Signature image is first preprocessed before feature extraction. The image is cropped to the bounding rectangle frame of the signature. This frame touches the signature at four different directions: left, right, top and bottom. So that only signature has been extracted from the total image. Figure 2 shows cropped signature image

Step 2 Maximum length and width of all person's signatures are determined. To overcome image scaling each handwritten signature image is resized

4 Wavelet Transform and Feature Extraction Algorithm

4.1 Wavelet Transform

Wavelet Transform is powerful mathematical tool for analyzing an image at several levels of resolution. The transform of a signal is just another form of representing the signal. It does not change the information content present in the signal. The Wavelet Transform provides a time-frequency representation of the signal. Wavelet Transform uses multi-resolution technique by which different frequencies are analyzed with different resolutions. The Wavelet Transform, at high frequencies, gives good time resolution and poor frequency resolution, while at low frequencies; the Wavelet Transform gives good frequency resolution and poor time resolution. Wavelet transform decomposes a signal into high frequency and low frequency components. By applying wavelet transform wavelet coefficient of signal are extracted. In the two-band wavelet transform, signal can be expressed by wavelet and scaling basis functions at different scale, in a hierarchical manner. For two-dimensional wavelet transform, an image is represented in terms of translations and dilations of a scaling function and a wavelet functions which can

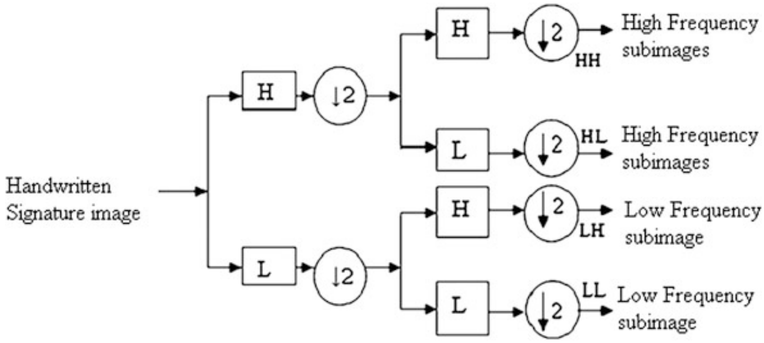


Fig. 3 Wavelet decomposition

be computed using a 2D filter bank consisting of low-pass and high-pass filters. After 2D decomposition, the given image is decomposed into several frequency components at multiple levels of resolution as shown in Fig. 3.

The sub bands HHk, HLk, LHk, k = 1,…, J are called the details, where k is the scale, with J being the largest scale in the decomposition, and a subband. The low pass information represents a smoothed version of original image and the main body of the original data. The high pass information represents sharper variations and details present in an image. For similar images, High frequency information represents difference in between images. A DWT with single decomposition levels has 3 + 1 frequency bands with 3 high-frequency bands [12]. Signature images undergo DWT analysis which results into image reduction.

4.2 Feature Extraction Algorithm

Signature images are decomposed up to second level of decomposition. After decomposition of images further high frequency wavelet coefficient are extracted. After performing first level of decomposition the high pass sub-images of signature these sub-images are divided to 5 * 5 blocks. Because of down sampling operation in DWT, sizes of sub-images of each level are different from next level (about 2 times). Due to that in the before reduction, these sub-images are resized to the nearest dimension of 5 multiples. Each 5*5 block of sub-image will be replaced by pixel. This causes further image reduction. The number of blocks of sub-images in each level resulting after portioning of sub images is different from other levels. Resizing is used just because of using 5 * 5 blocks in partitioning and does not have any effects in results. Further sub-image CH, CD, CV i.e., wavelet coefficients are coded. These sub-images are as shown in Fig. 4a, b, c.

Every 5 * 5 pixel block of high frequency sub-image is coded. In the second phase of reduction, a matrix is assigned to each sub-image that its dimension corresponds to the number of blocks of this sub-image. Sub-image matrix consists

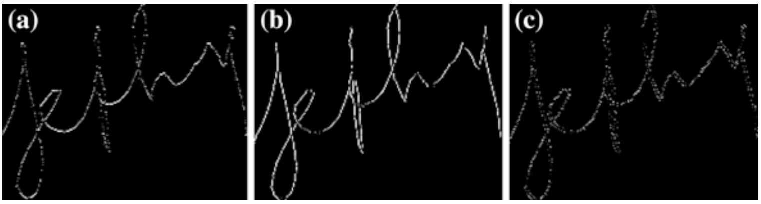


Fig. 4 High frequency coefficients of signature image

of coded elements either as ‘1’ or ‘0’. It consists of “1” if 5 % of its corresponded block pixels in sub-image are black unless ‘0’ is stored in. Coded High frequency sub images matrices are going to form feature vector.

5 Image Classification

Statistical classifiers are most common way for identification of template. The classification step in offline handwritten signature identification systems can be viewed as feature matching process between the features of a new handwritten signature and the features saved in the database. For successful classification, wavelet based significant features have been extracted and these set of feature vectors are saved in a database. Proposed algorithm uses hamming distance as a classifier. We have evaluated proposed algorithm using statistical classifiers such as Euclidean, minkowski, cosine, correlation, cityblock. Feature vector obtained from handwritten signature images comprises 1’s and 0’s. So hamming distance is used to calculate distance between training images and testing pattern. The Hamming distance (HD) between two Boolean vectors can be defined in equation 1 as follows [13]:

$$\text{Hamming Distance } = \frac{1}{N}\sum_{j=1}^{N} H_A(j) \oplus H_B(j) \tag{1}$$

Where, HA and HB are the coded wavelet coefficients of two signature images and N is the size of the feature vector. The $\oplus$ is the known Boolean operator that gives a binary 1 if the bits at position j in HA and HB are different and 0 if they are similar. After performing evaluation of algorithm using these different classifiers we have got similar accuracy. There is no significant improvement in accuracy.

6 Experiment and Result

In this section, the proposed method is analyzed and evaluated by experiments with the Caltech university database [14] and the UCOER database respectively. In order to test our proposed offline signature recognition algorithm, we have

Fig. 5 UCOER database sample signature images

conducted experiments on two signature databases. The Caltech datasets used for testing of the proposed approach have been obtained by the use of a camera based interface for signature tracking [14]. The data is organized in two sets: Set 1 consists of the signatures of 56 individuals with 25 genuine and 9 skilled forged signatures of each subject; Set 2 consists of signatures of 50 different individuals with 30 signatures of each subject. The proposed system was tested on dataset-1. In this experiment, sixteen images per person are randomly chosen as the training image set, and the other images are used as the testing image. Proposed algorithm gives recognition accuracy of 95 %. We have also determined FAR and FRR for our proposed algorithm. FAR and FRR of our offline signature recognition system is 2.67 and 0.22 resp. Also we have got genuine accept rate of our proposed system as 97.32. Proposed algorithm was also evaluated on forgery database. By selection of proper threshold we have got FAR of 1.19 for forgery images. We have carried out one more experiment on UCOER database of 150 signatures collected from 30 individuals which consists of 5 signatures from each individual. Sample database images are as shown in Fig. 5. The signatures were collected using either black or blue ink on a white sheet of paper. A scanner subsequently digitized the signatures contained on each page with a 300 dpi resolution in 256 grey-levels. Afterwards the images were cut and pasted and were saved separately in files. In this experiment we have used 4 random signature images for training and 1 image for testing. Experimentally, we have seen that the accuracy on this database of the proposed system is about 90.00 %. Figs. 6, 8 shows recognition accuracy graph on Caltech database and UCOER database along with variation in number of training

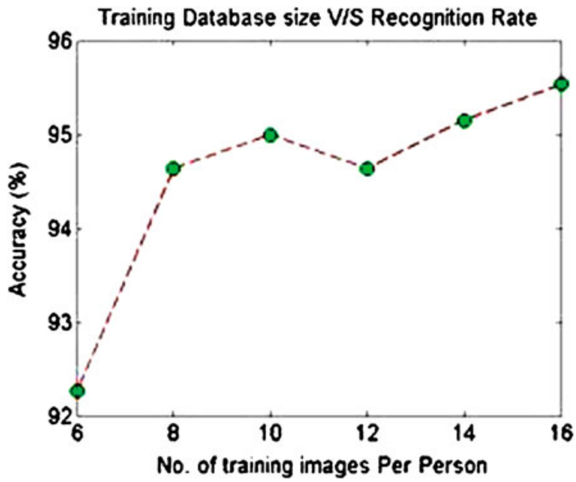


Fig. 6 Recognition accuracy (caltech)

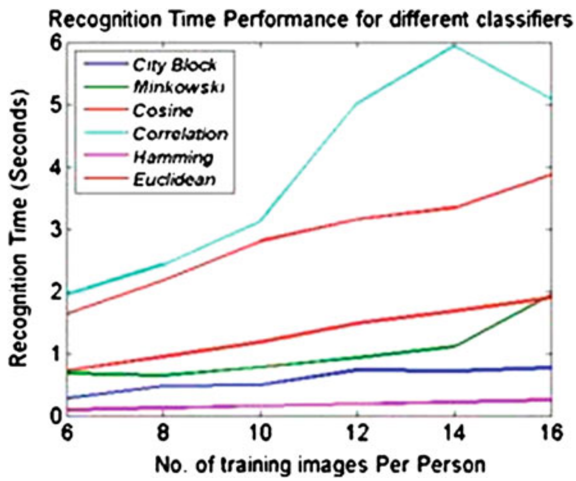


Fig. 7 Recognition time (caltech)

images. Fig. 7, 9 represents graph for recognition time for statistical classifiers along with variation in number of training images. For Hamming distance as a classifier we have required minimum recognition time as 0.2558 s on Caltech database with 16 training images and 0.0241 s on UCOER database with 4 training images. Hamming distance performs faster signature identification as compared to other classifiers. Proposed algorithm recognizes human signature within very less time.

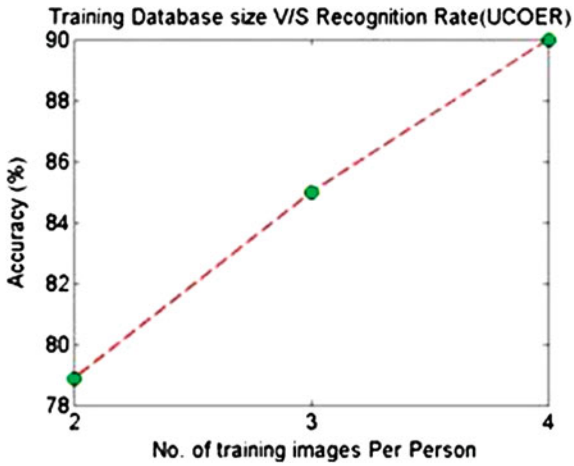


Fig. 8 Recognition Accuracy (UCOER)

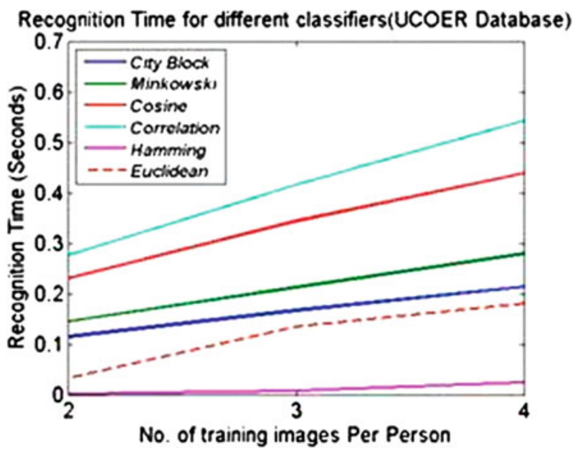


Fig. 9 Recognition time (UCOER)

7 Conclusion

In this paper, a novel offline handwritten signature recognition system is presented. A novel 2D-DWT signature feature extraction scheme is proposed for signature recognition. The feature set of a signature, is supposed to represent global attributes of the signature whereas signature is considered purely as a image. In proposed method we have used second level multi-resolution dwt features. We have performed recognition accuracy and recognition time analysis for different statistical classifiers. During the experiment carried it was observed that hamming distance as a classifier outperforms other statistical classifiers. So We have used hamming distance based algorithm for classification of images. We have achieved maximum success rate of our algorithm as 95 % and FAR as 0.22. Proposed methodology has given good results for FAR. Also proposed algorithm identifies signature within very less time. Hamming classifier requires less time for authorizing human signature as compared to other classifiers. During this evaluation it has been observed that as signature images are random in nature. This randomness property can affect accuracy of system. In the future, other biometric features can be combined with signature features to boost system's recognition rate and to decrease its error rate.

References

1. Jain AK, Ross A, Prabhakar S (2004) An introduction to biometric recognition IEEE Trans. Circuits Syst. Video Technol 14 (1):4–20
2. Bowyer K, Govindaraju V, Ratha N (2007) Introduction to the special issue on recent advances in biometric systems. IEEE Trans Syst Man Cybern—B 37(5):1091–1095
3. Franke K, delSolar JR, open MK̈ (2003) Soft-biometrics: soft computing for biometric-applications. Tech Rep IPK
4. Impedovo S, Pirlo G, (2007) Verification of handwritten signatures: an overview. In: ICIAP'07: Proceedings of the14th international conference on image analysis and processing, IEEE computer society. Washington, USA, pp 191–196doi.org/10.1109/ICIAP.2007.131
5. Madasu VK, Lovell BC, Kubik K (2005) Automatic handwritten signature verification system for Australian passports. In: Science, engineering and technology summit on counterterrorism technology, Canberra, 14 July p 53–66
6. Peter SD, Hong-Yuan ML (1999) "Wavelet-Based Off-Line Handwritten Signature Verification". In: Computer vision and image understanding, pp 173–190
7. Samanesh G, Mohsen EM (2009) "Off-line persian signature identification and verification based on image registration and fusion". In: J Multi 4:137–144
8. Larkins R, Mayo M (2008) "Adaptive feature thresholding for off-line signature verification". In: Image and vision computing New Zealand pp 1–6
9. Fakhlai M, Pourreza H (2008) Off line signature recognition based on wavelet, curvelet and contourlet transforms. 8th WSEAS international conference on signal processing and computational geometry and artificial vision (ISCGAV"08), Rhodes, Greece, Aug pp 20–22
10. Prakash HN Guru DS (2010) Offline signature verification—an approach based on score level fusion. Int J Comput Appli 1(18):0975–8887
11. Vargas JF, Ferrer MA, Travieso CM, Alonso JB (2011) Off-line signature verification based on grey level information using texture features. Pattern Recogn 44:375–385)
12. Gonzalo P, Jesús Manuel de la C (2004) "A wavelet-based image fusion tutorial", Pattern Recognition, Elsevier Science Inc, Sep 37(9):1855–1872
13. Daugman CJ (1998) Recognizing Persons by their Iris Patterns, in Biometric. In: Jain A, Bolle R, Pankati S (eds) Personal Identification in Networked Society, Kluwer, 103–121
14. http://www.vision.caltech.edu/mariomu/research.html
15. Zhang D, Campbell J, Maltoni D, Bolle R (2005) Special issue on biometric systems. IEEE Trans Systems, Man and Cybern—C 35(3):273–275
16. Prabhakar S, Kittler J, Maltoni D, Gorman L.O', T.Tan (2007) Introduction to the special issue on biometrics: progress and directions PAMI29 (4):513–516

Hardware Overhead vs. Performance of Matrix Multiplication on FPGA

Ju Seong Lee, Sang Don Kim, Yeong Seob Jeong and Seung Eun Lee

Abstract Matrix multiplication requires a large number of operations, demanding for high performance computing. In order to complete the matrix multiplication in one clock cycle, a designer can utilize multiple multipliers. However, this approach is inefficient in terms of hardware area and power consumption. Therefore, it is important to find out the way to complete the multiplication that is fast and uses hardware resources properly. In this paper, we introduce the way to reduce the number of multipliers and provide the hardware overhead and performance of matrix multiplication on FPGA.

Keywords Matrix multiplication · Digital signal processing · Low-power design · FPGA

1 Introduction

The latest advanced systems show the outstanding performance and the excellent single processing capabilities. Because of such the outstanding performance, these systems need a large amount of multipliers in matrix multiplication for the much operation. Some mobile systems that use these modern devices also have spatial constraints in hardware area since they should cover low power consumption and the ability to execute multiple applications.

J. S. Lee · S. D. Kim · Y. S. Jeong · S. E. Lee (✉)
Department of Electronic IT Media Engineering, Seoul National University of Science and Technology, Seoul, Korea
e-mail: seung.lee@seoultech.ac.kr

Mohan S. and S. S. Kumar (eds.), *Proceedings of the Fourth International Conference on Signal and Image Processing 2012 (ICSIP 2012)*, Lecture Notes in Electrical Engineering 222,
DOI: 10.1007/978-81-322-1000-9_28,

Most of all, multiplication needs more computation time and hardware area particularly than addition in hardware implementation. Thus, effort to reduce the number of multipliers has been conducted from the past until now. Volker present an algorithm in order to compute the coefficients of the product between invertible two square matrices [1]. Matrix multiplication has been applied in signal processing and various field. In particular, such as image processing computation of the matrix has a significant influence on performance of whole system when a large amount of data must be processed. Therefore, there were studies to deal with enormous data and find effective methods in the matrix multiplication [2, 3]. Colin proposed an analytical model for estimation of the performance of FPGA-based dense and sparse matrix multiplication performance [4]. In case of parallel matrix multiplication, it requires a large amount of multipliers for much matrix multiplication [5]. Area of device and computation time can be changed by computation method and order on the matrix multiplication computation that has three or more different sizes.

In this paper, we introduce the way to reduce the number of multiplications and provide the size of a device and the difference of computation time in accordance with various computation methods in case of parallel matrix multiplication. We analyze the operation process, required resource and processing time between a normal and the associative law applied computation in Sect. 2. In the Sect. 3, the required hardware is implemented based on our proposal. The experimental result is shown in Sects. 4 and 5 concludes the paper.

2 Design Diversity Metric and Reliability Analysis

When the matrix multiplication for A with the size of $(n \times m)$ and B with the size of $(m \times p)$, multiplication is needed m times in each components. If there are three or more matrixes, multiplication is required much more. In this case, we can use Associative law established by calculation law. It is able to minimize the number of multiplication. The matrixes A, B and C are assumed with the different size of $A(n \times m)$, $B(m \times p)$, $C(p \times q)$.

$$A = \begin{pmatrix} A(1 \times 1) \cdots A(1 \times m) \\ \vdots \quad \cdots \quad \vdots \\ A(n \times 1) \cdots A(n \times m) \end{pmatrix}, B = \begin{pmatrix} B(1 \times 1) \cdots B(1 \times p) \\ \vdots \quad \cdots \quad \vdots \\ B(m \times 1) \cdots B(m \times p) \end{pmatrix}$$
$$C = \begin{pmatrix} C(1 \times 1) \cdots C(1 \times q) \\ \vdots \quad \cdots \quad \vdots \\ C(p \times 1) \cdots C(p \times q) \end{pmatrix} \tag{1}$$

Figure 1 shows the three ways of multiplication. In Case 1, multiply three matrixes at a time as a base case. In Case 2 and 3, apply the Associative law to matrix multiplication. If multiply all the components at a time, the parallel operation is maximum. It takes only one clock cycle. However, the number of multiplier is exponentially increased by proportional to the size of matrix. It can cause problems in power consumption and production cost. Thus, it is necessary to find the optimal method that is faster and has high performance.

When multiply the three matrixes, calculating the result matrix's components one by one in one clock cycle. For the faster calculation, we have to decide the number of the parallel operation. In case that multiply $A(n \times m)$ by $B(m \times p)$, multiplication is needed m times in one clock. Similarly, in case that multiply $B(m \times p)$ by $C(p \times q)$, multiplication is needed p times. This method cannot do 'Three or more matrixes multiplication' in one clock. But, it uses multipliers less than case 1. According to position applied the Associative law, the number of multiplications can be changed such as Case 2, Case 3. Therefore, we can find the better way that uses hardware area and have fast computation time.

Figure 2 shows block diagrams of case 1, case 2 and case 3. The case 1 represents the $A{\cdot}B{\cdot}C$ multiplication. This operation is completed at a time. But it requires actually more space because it needs lots of combinational logics. The case 2 and case 3 show operations of the $(A{\cdot}B){\cdot}C$ and $A{\cdot}(B{\cdot}C)$ multiplication based on the associative law. This operation requires multiple clock cycles and each of the matrix elements is calculated per clock cycle. Therefore, the number of multipliers is determined with the large number either m or p. And the multiplication is calculated according to the associative law. $(n \times m){\cdot}(m \times p)$ calculations are performed first in case 2. At this point, the parallel processing is carried out with m number of multipliers. The operation needs clock cycle as much as $(n \times p)$. As a result of the operation $(n \times p)$ matrix comes out. Then $(n \times p){\cdot}(p \times q)$ operation is performed. In this case, the parallel processing is carried out with p number of multipliers. And the operation needs clock cycle as much as $(n \times q)$.

The total clock cycle for multiplication is $(n \times p) + (n \times q)$. $(m \times p){\cdot}(p \times q)$ calculations are performed at first in case 3. At this point, the parallel processing is carried out with p number of multipliers. The operation

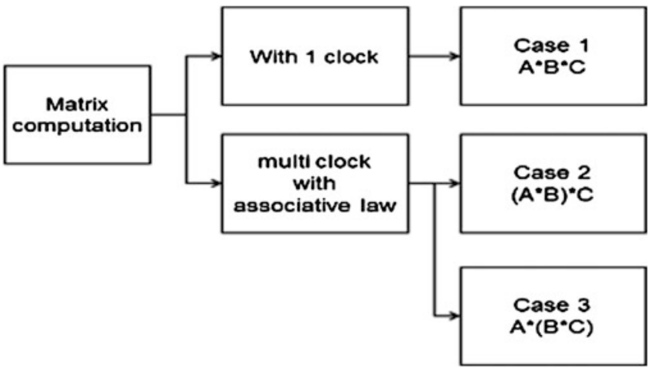


Fig. 1 Matrix multiplication configuration

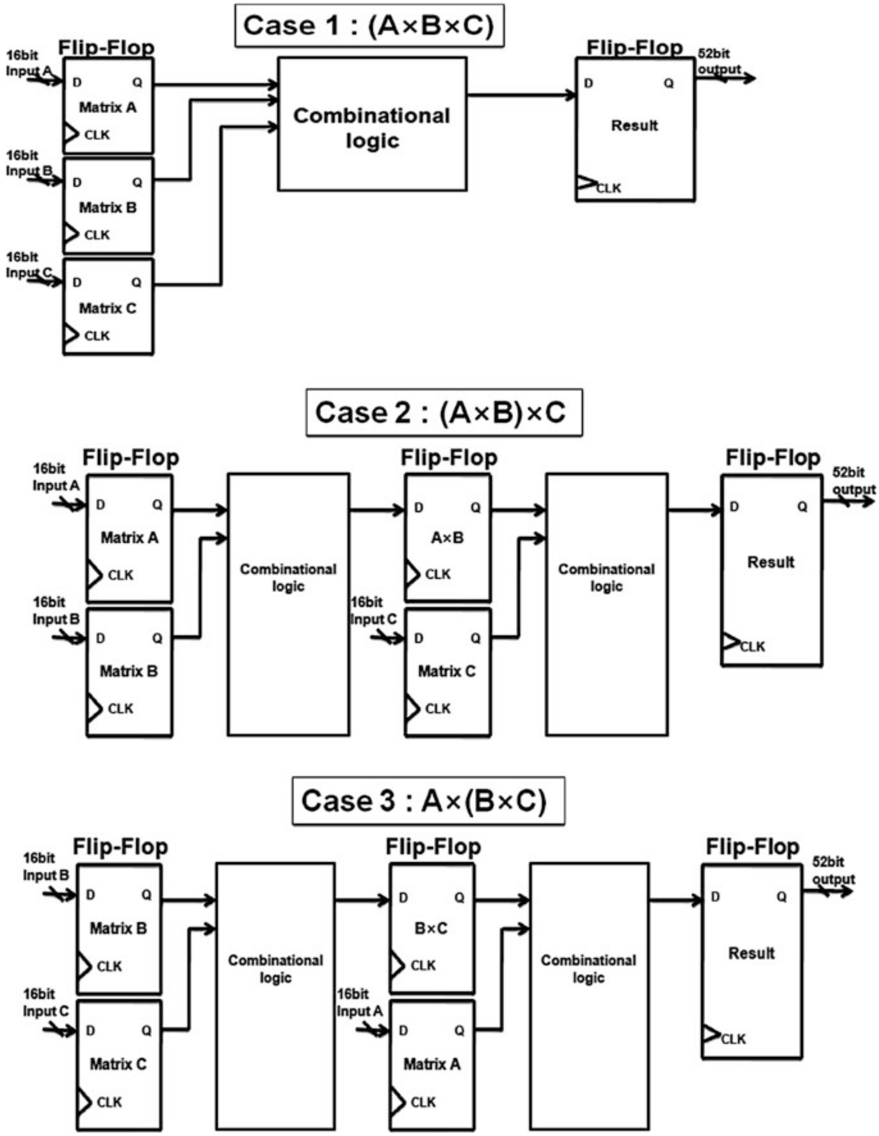


Fig. 2 Block diagram of each case

needs clock cycle as much as $(m \times q)$. And then $(n \times m)\cdot(m \times q)$ operation is performed. The parallel processing is carried out with m number of multipliers in this case. The operation needs clock cycle as much as $(n \times q)$. So the total clock cycle is $(m \times q) + (n \times q)$. For example, suppose there are $A(2 \times 3)$, $B(3 \times 8)$, $C(8 \times 2)$ matrixes. First, apply the case 1 to $A \times B \times C$. It can be finished in only one clock cycle, and the number of multiplication is 192. Required multiplier in

$A \times B$ is three. And required multiplier in $B \times C$ is eight. According to this, the number of required multiplier is eight in case 2 and 3. The number of multiplications when compute parallel in Case 2, Case 3 are as follows:

$$\text{Case 2}: \ (\mathrm{A} \times \mathrm{B}) \times \mathrm{C} \ = 2 \times (3/3) \times 8 + 2 \times (8/8) \times 2 = 16 + 4 = 20 \quad (2)$$

$$\text{Case 3}: \ \mathrm{A} \times (\mathrm{B} \times \mathrm{C}) \ = 3 \times (8/8) \times 2 + 2 \times (3/3) \times 2 = 6 + 4 = 10 \quad (3)$$

Case 2 requires 20 times multiplications. And case 3 requires 10 times multiplications. As the result, the number of multipliers and required clock cycles are determined by the operation order. Therefore the number of calculation of the case 3 is more than case 2 when m is larger than p. On the other hand, the number of calculation of the case 2 is more than case 3 when p is larger than m.

3 Implementation

The Xilinx ISE 11.1 tool is used in the synthesis & implementation and we exploit the spartan3 XC3S4000-FG676 device under −5 speed grade. For the simulation, we use the modelsim student edition 10.1b of the Mentor graphics co. The purpose of this paper is to find the optimized number of multipliers from those of different matrix based on the Associative law. First of all, we classify into three cases that are case 1: calculation completed within 1 clock cycle, case 2: $(A{\cdot}B){\cdot}C$ and case 3: $A{\cdot}(B{\cdot}C)$. They are based on A, B, and C matrix mentioned Sect. 2. We have implementation and simulation with the three types.

Figure 3 is the simulation result for the case 1. The maximum clock cycle is 10 MHz. When entering input, output comes out after 1 clock cycle. Figure 4 is the simulation result for the case 2. The maximum clock cycle is 62 MHz. When entering input, output comes out after 20 clock cycle. Figure 5 is the simulation result for the case 3. The maximum clock cycle is 63 MHz. When entering input, output comes out after 10 clock cycle. Thus, it was demonstrated correctly for the operation. It can also be confirmed as calculated result in Sect. 2.

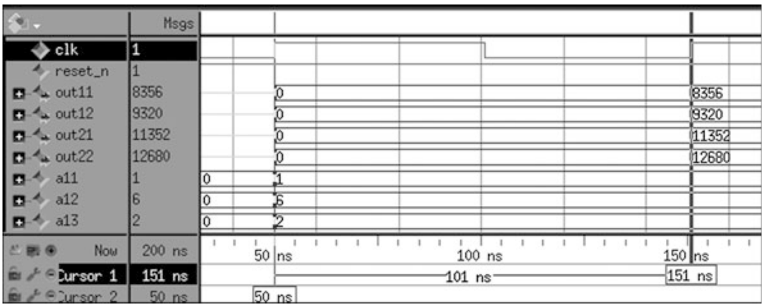


Fig. 3 Simulation result of the case 1

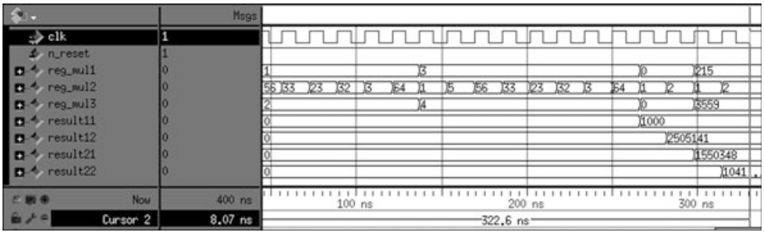


Fig. 4 Simulation result of the case 2

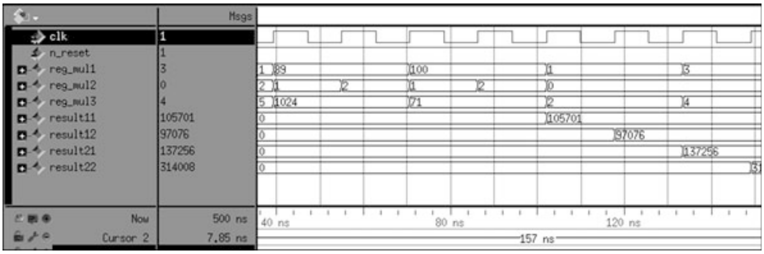
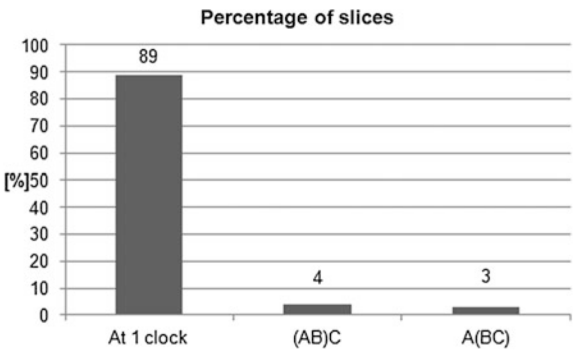


Fig. 5 Simulation result of the case 3

4 Experimental Result

Figure 6 shows the number of multipliers and the percentage of slices after implementation in device. The efforts to reduce the chip cost are also applied. The number of multipliers is 144 in case 1 and the eight is used in case 2 and case 3. Slices of 24,613 are used in case 1. And total numbers of slices are 27,648 (using 89 %). This result shows that a very large area was used. Slices of 1,198 and slices of 1,019 are used in each case 2, 3. These are relatively a small area (using in each case 4 and 3 %). Figure 7 shows the time which takes to complete multiplication each case. 101, 322.6 and 157 nano second was consumed in each case. The results

Fig. 6 The area cost of the matrix multipliers in terms of percentage of slices



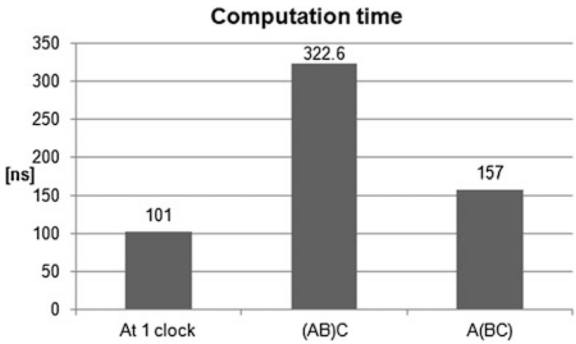


Fig. 7 The computation time for the matrix multipliers

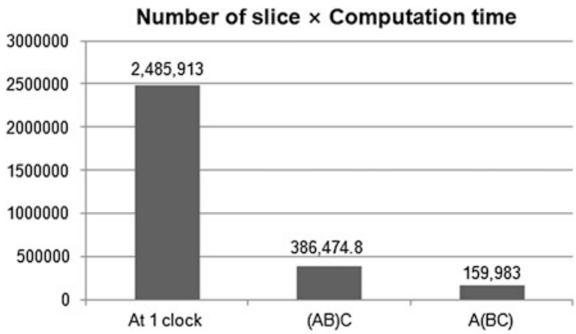


Fig. 8 Performance of the matrix multipliers in terms of area and computing time

of number of slices × Computation time are shown in Fig. 8. It can help that confirming the efficiency of Case 2, 3.

$$\text{Case } 1: A \cdot B \cdot C = 101\,\text{ns} \times 1\text{clock cycle} \;= 101\,\text{ns} \tag{4}$$

$$\text{Case } 2: A \cdot (B \cdot C) \;= 16.13\,\text{ns} \times 20\,\text{clock cycle} \;= 322.6\,\text{ns} \tag{5}$$

$$\text{Case } 3: \;(A \cdot B) \cdot C = 15.7\,\text{ns} \times 10\,\text{clock cycle} \;= 157\,\text{ns} \tag{6}$$

According to result, case 1 is the fastest way to complete the multiplication. But it requires plenty of hardware area. Otherwise, case 2 requires a small hardware area. But it has slow operation speed. And, case 3 using associative law not only requires a small hardware area but has proper operation speed.

5 Conclusion

In this paper we analyzed the number of the required multiplication, the time to complete the operation, the number of gate used in hardware implementation and maximum calculation time when the combine of three or more matrix that have different sizes changes based on associative law. As a result, a large amount of device gates were demanded in order that all multiplication operations had finished in 1 clock cycle. The maximum frequency that can be operated is low because of many combinational logic and long critical path. On the other hand, the number of gate used in hardware implementation is greatly reduced, critical path is relatively small and the maximum operating frequency is quite high when we found an appropriate combination of the matrix based on associative law. In this experimental result, there is much difference in the number of calculation to finishing whole processing according to the sequence of matrix calculation. When using proper processing sequence, the computation time is similar to the fastest way to complete the multiplication with 1 clock cycle.

Acknowledgments This study was supported by Seoul National University of Science and Technology, Korea.

References

1. Strassen V (1969) Gaussian elimination is not optimal. Numer Math 13:354–356
2. Choi J (1997) A fast scalable universal matrix multiplication algorithm on distributed-memory concurrent computers. In: Proceedings of the 11th international symposium on parallel processing
3. Bensaali F, Amira A, Bouridane A (2005) Accelerating matrix product on reconfigurable hardware for image processing applications, circuits, devices and systems. In: IEEE Proceedings 3 June 2005
4. Lin CY (2011) A model for matrix multiplication performance on FPGAs. In: 21st international conference on field programmable logic and applications
5. Al-Qadi Z, Aqel M (2009) Performance analysis of parallel matrix multiplication algorithms used in image processing. World Appl Sci J

SDRAM Controller for Retention Time Analysis in Low Power Signal Processor

Sang Don Kim, Yeong Seob Jeong, Ju Seong Lee and Seung Eun Lee

Abstract The SDRAM requires the refreshing at every refresh time to maintain the data, and this operation consumes power. Because the power consumption of the processor is decreased, the power cosnumption on the SDRAM is taking large portion of the total power consumption. The refresh time can be expanded because the retention time and power consumption can be changed. In this paper we introduce the SDRAM controller which enables the analysis of the retention time for the power redectoin purpose.

Keywords Low-power design · Resilient design · SDRAM controller

1 Introduction

The SDRAM is important in signal processing because of its fast speed and sufficient capacity. The processor power consumption is decreasing according to the development of semiconductor facilities. The SDRAM requires the refresh in order to retain the data because it will be discharged over time. The refresh time is relevant with the retention time that can be retaining the data on SDRAM in the presence of leakage current. There are studies in the retention time on SDRAM have been conducted to reduce refresh overhead. The study of the retention time is important concern in the past though different manufacturing processes is used in the present [1]. Lee et al. proposed the DRAM cell with metallic shield embedded

S. D. Kim · Y. S. Jeong · J. S. Lee · S. E. Lee (✉)
Department of Electronic IT Media Engineering, Seoul National University of Science and Technology, Seoul, Korea
e-mail: seung.lee@seoultech.ac.kr

Mohan S. and S. S. Kumar (eds.), *Proceedings of the Fourth International Conference on Signal and Image Processing 2012 (ICSIP 2012)*, Lecture Notes in Electrical Engineering 222,
DOI: 10.1007/978-81-322-1000-9_29,

for improvement of retention time [2]. The research on the retention time for the individual cell of DRAM has studied [3]. Cho et al. proposed a convex channel structure for the capacitor-less DRAM Generation 2 type cell to form a physical well for holes storage and to improve retention time [4]. The convex channel structures show higher potential electrostatic values since beneath the gate stores more holes. Liu proposed a low-cost mechanism that can identify and skip unnecessary refreshes by exploiting inter-cell variation in leakage current of capacitor and transistor PN-junction decrease [5].

Therefore, the retention time can be changed and the refresh time can be extended if retention time is longer than refresh time. In that case, we can decrease the power consumption of the SDRAM refresh, reducing the total power consumption of system. In addition, we can activate the more other command on the SDRAM because the NOP (No Operation) time is extended.

2 The Characteristic of the SDRAM in Retaining the Data

The SDRAM retention time is the time that the data is stabling after write the data to the SDRAM. Therefore, the "READ" command is required after the "WRITE" command and a certain time for measuring the new retention time. The followings are considerable factors to measure the retention time. The SDRAM has matrix structure that is classified into row address and column address. The refresh command is active at each row address. Commonly used refresh time is 64 *ms* and all rows should be refreshed during this time. The SDRAM cannot receive any other commands during this time to activate refresh command. The SDRAM controller can exploit useful methods for activating refresh about all rows. That is the "Distributed Refresh" and the "Burst Refresh". The "Distributed Refresh" has a same term between the each refresh command during the refresh time while the "Burst Refresh" refreshes all rows after the refresh time. To conduct the "WRITE" or the "READ" command operation, the SDRAM refresh command should not be working. The "READ" and the "WRITE" commands are contained within row line access. To access same row continuously can bring an error on measuring of the retention time because of the SDRAM structure. The precharge operation is required to activate a "READ" or a "WRITE" command. SDRAM command is long when precharge is separately activated. For the more short command, SDRAM supports interleaved mode. The SDRAM has two addresses; it is the row address and the column address. The interleaved mode has the SDRAM automatically precharging between the inputs of the two addresses. Therefore, the command with auto-precharge is shorter than separated precharge command. The SDRAM provides burst mode for the fast operation. When the "READ" or the "WRITE" command is activated, data can be accessed the pre-set burst length. The burst mode is faster than single location mode but it is accessed in same row. Therefore this is unsuitable when measuring the retention time of the SDRAM. In this paper we set to burst length to 1.

3 Implementation

3.1 Initialize and Load Mode Register

The SDRAM requires the initialization sequence after the power-up for processing the other commands (See Fig. 1). In this paper, the SDRAM controller disabling refresh operation is implemented on FPGA. We adopted the Cyclone IV (EP4CE115F29C7N) FPGA and the 64 MB SDRAM (ISSI), they are integrated on a DE2-115 Development Board. The SDRAM controller sequentially processes the following commands for the initialization: 1) controller wait during the 200*us* when SDRAM is powered-up and clock is stable; 2) precharge the SDRAM and activate the auto refresh at least 8 times; 3) decides the SDRAM operating condition using the LMR(Load Mode Registers) like as the Write Burst Mode, the CAS Latency, the Sequential/Interleaved Mode and the Burst Length. The burst length is set to 1 and it processes 1 data at a time. SDRAM initial sequence can vary depending on the type of the SDRAM. Other command can be operated when the initialization sequence is ended.

3.2 Single Write with Auto Precharge

SDRAM has row address and column address; these two addresses are required for the write data to the SDRAM. Firstly, SDRAM controller opens the row line using the *RAS* signal. And wait 1 clock cycle. In this waiting cycle, auto-precharge is activated in the SDRAM. Next, it writes a data using the *CAS* and the *WE* signals. The *DQM* (Data input/output mask) has to low level in order that the SDRAM data register can be changed. The write command is required 3 clocks for a data write

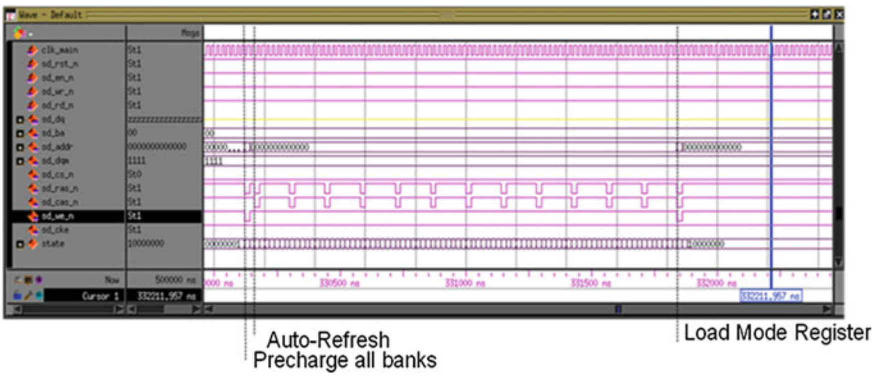


Fig. 1 SDRAM initialization

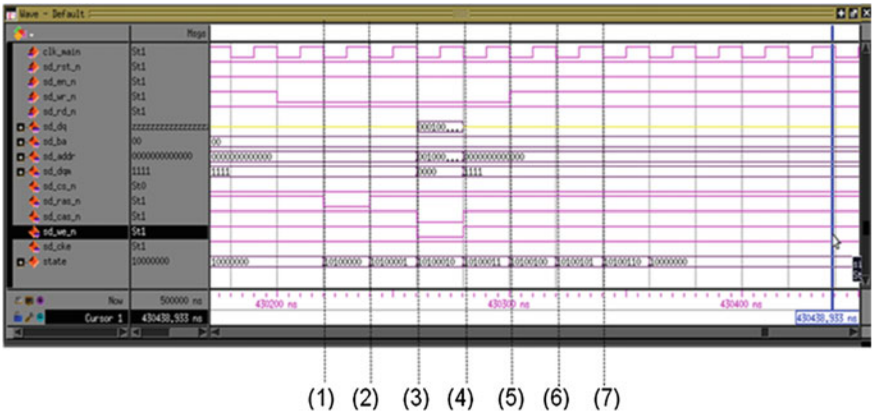


Fig. 2 SDRAM write operation

(See Fig. 2). However it consumes 7 clocks in accordance with the interval between the active and the active.

3.3 Single Read with Auto Precharge

SDRAM "READ" command is operating in the same way as the "WRITE" command except for the *WE* signal. However, sensing the data which is stored in the capacitor of the SDRAM is a physically time-consuming task. Therefore, the actual time delay on reading the data will occur after a "READ" command. This is represented by CAS Latency. SDRAM can read valid data after the 2 or 3 clocks and it can be pre-programmed in the initialization process. If the CAS latency is

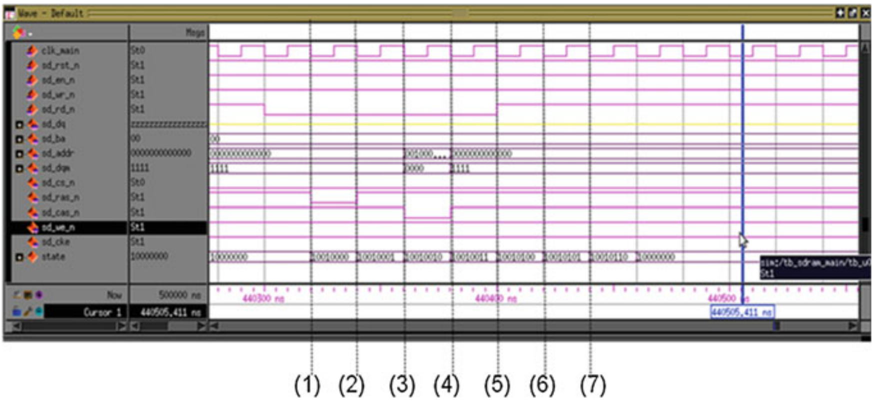


Fig. 3 SDRAM read operation

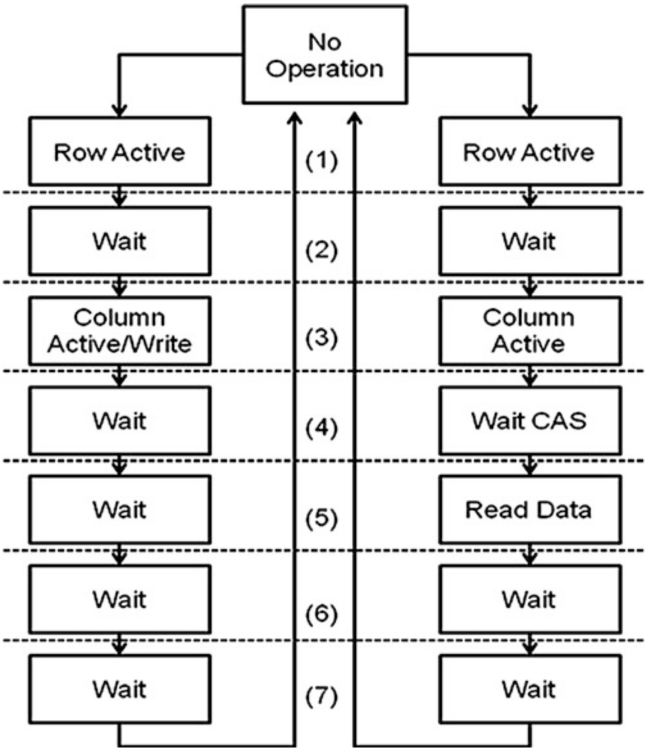


Fig. 4 SDRAM write/read flow

set to 3, the time that valid data is coming out takes longer than 2 CAS Latency. The "READ" operation will consume maximum 6 clocks in accordance with CAS latency when burst length is 1 (See Fig. 3). But it will consume 7 clocks in accordance with the interval between the active and the active. In this paper, this CAS latency does not affect the length of the command in read operation. Figure 4 shows operating sequence of the SDRAM controller "READ" and "WRITE" commands on 2 CAS Latency, it requires 7 clock for each command.

4 Experimental Result

The SDRAM can be controlled with our controller in FPGA. It is possible to confirm the normal operation with the 50 MHz main clock. The maximum operating frequency is 100 MHz in case of 2 CAS latency with FPGA using the internal PLL. The "WRITE" operation is to write the dedicated data to SDRAM. And the "READ" operation is to read the data from SDRAM after the certain time. This controller does not operating the refresh command. Figure 5 shows the data written to the SDRAM using the implemented controller. The written data is

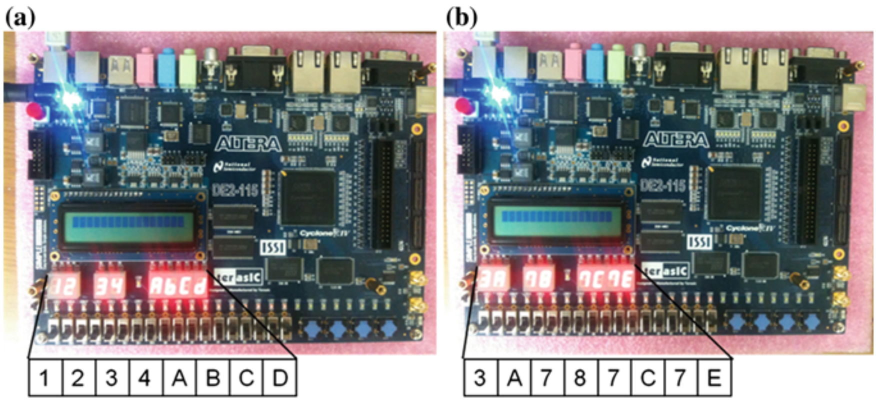


Fig. 5 Destroyed the data over the time without the refresh command

"1234ABCD" as hex digit, and seen as 7-segment. The SDRAM is no data loss during 64 *ms* after the write operation because SDRAM has retention time (See Fig. 5 (a)). This data is destroyed to "3A787C7E" after the certain time; 7-segment shows this result (See Fig. 5 (b)).

5 Conclusion

The power consumption of the SDRAM block can be decreased when the proper refresh time is adopted. Therefore, measuring the retention time is important to decide the proper refresh time. In this paper we implemented a SDRAM controller disabling the refresh command in order to measuring the retention time of the SDRAM. The low power SDRAM controller is very suitable for the low power system. In addition The SDRAM's throughput also increases when the refresh operation time decreases. Therefore, it is important to expand the refresh time. We plan to analysis the retention time of SDRAM by using our SDRAM controller, targeting low power embedded system designs.

Acknowledgments This study was supported by Seoul National University of Science and Technology.

References

1. Hamamoto T (1998) On the retention time distribution of dynamic random access memory (DRAM). Electron Devices 45(6):1300–1309
2. Lee SH, Hong SH, Oh JH, Choi YK, Bae DI, Park SH, Roh BH, Chung TY, Kim K (2003) Improvement of data retention time using DRAM cell with metallic shield embedded (MSE)-STI for 90 nm technology node and beyond. Eur Solid-State Device Res 151–154

3. Weber A, Birner A, Krautschneider W (2005) Data retention analysis on individual cells of 256 Mb DRAM in 110 nm technology. Solid-State Device Research Conference, pp 185–188
4. Cho MH, Shin C, Liu TJK (2009) Convex channel design for improved capacitorless DRAM Retention Time. Simul Semicond Process Devices 1–4
5. Liu J (2012) RAIDR: retention-aware intelligent DRAM refresh, Computer architecture (ISCA). International Symposium onComputer Architecture (ISCA), pp 1–12

Expert System Design Based on Wavelet Transform and Non-Linear Feature Selection

Niladri Prasad Mohanty and Pritish Ranjan Pal

Abstract Electromyogram (EMG) is the record of the electrical excitation of the skeletal muscles, which is initiated and regulated by the central, and peripheral nervous system. EMGs have non-stationary properties. EMG signals of isometric contraction for two different abnormalities namely ALS (Amyotrophic Lateral Sclerosis) which is coming under Neuropathy and Myopathy. Neuropathy relates to the degeneration of neural impulse whereas myopathy relates to the degeneration of muscle fibers. There are two issues in the classification of EMG signals. In EMG's diseases recognition, the first and the most important step is feature extraction. In this paper, we have selected Symlet of order five of mother wavelet for EMG signal analysis and later six non-linear features have been used to classify using Support Vector Machine. After feature extraction, feature matrix is normalized in order to have features in a same range. Simply, linear SVM classifier was trained by the train–train data and then used for classifying the train-test data. From the experimental results, Lyapunov exponent and Hurst exponent is the best feature with higher accuracy comparing with the other features, whereas features like Capacity Dimension, Correlation Function, Correlation Dimension, Probability Distribution & Correlation Matrix are useful augmenting features.

Keywords ALS · Hurst exponent · Lyapunov exponent · Myopathy

N. P. Mohanty (✉)
Department of Biomedical Engineering, National Institute of Technology, Raipur, Chhattisgarh, India
e-mail: niladree.mohanty@gmail.com

P. R. Pal
Department of Electrical Engineering, SGSITS, Indore, MP, India
e-mail: pritishpal.pal@ymail.com

Mohan S. and S. S. Kumar (eds.), *Proceedings of the Fourth International Conference on Signal and Image Processing 2012 (ICSIP 2012)*, Lecture Notes in Electrical Engineering 222,
DOI: 10.1007/978-81-322-1000-9_30,

1 Introduction

Movement and position of the limbs are controlled by the electrical signals travelling forward and backward between Muscle fibers, Peripheral and Central Nervous System [1, 2]. Conscientious Registration and interpretation of these muscle electrical potential is called as Electromyogram (EMG). Figure 1 shows EMG Signal of Normal, ALS, Myopathy patient from Biceps Brachii region.

Due to the emanation of Pathological condition in motor system, whether in spinal cord, the motor neuron, the muscle or the neuromuscular junction the characters of electrical potentials generated during the contraction and relaxation of muscles changes [4]. Careful registration and study of electrical signals in muscles thus can be valuable aid in discovering and diagnosis abnormalities not only in muscles but also in the motor system as a whole [3, 5].

EMG classification is one of the most difficult pattern recognition problems because there usually exists small but numerous variations in EMG features, which leads to difficulty in analyzing EMG signals. In muscles diseases recognition, there are two main points, namely *feature selection* and *classifier design*. In general, the methods of feature selection can be divided into two types: the measure of classification accuracy and the valuation using statistical criterion. After that the selection of the best features based on the proposed statistical criterion method is investigated.

For this purpose, we evaluate different kinds of features that have been widely used in EMG diseases recognition. The results of this evaluation and the proposed statistic method can be widely used in EMG applications such as control of EMG robots and prostheses or the EMG diagnosis of nerve and muscle diseases [6–8].

The sects. 2 and 3 explain about the data selection wavelet analysis, and the features extracted. The sect. 4 gives the results, discussion, and conclusion of the context.

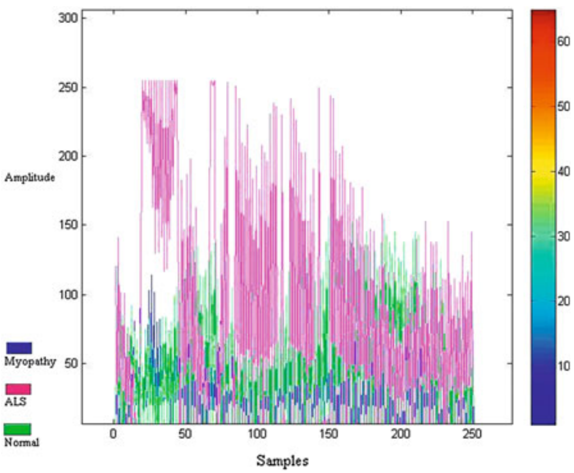


Fig. 1 EMG Signal of Normal, ALS, Myopathy patient from Biceps Brachii region

2 Data Selection

2.1 Subjects

For our study, we have chosen 3 groups. First group consisted of 10 normal subjects aged 21–35 years, 4 females and 6 males. All of the 10 subjects are in very good physical condition except one. None in this group had signs or history of neuromuscular disorders. The group with myopathy consisted of 7 patients; 2 females and 5 males aged 19–63 years. The ALS group consisted of 8 subjects; 4 females and 4 males aged 35–67 years. Besides clinical and electrophysiological signs compatible with ALS, 5 of them died within a few years after onset of the disorder, supporting the diagnosis of ALS.

2.2 Experimental Protocol

The EMG signals were recorded under usual conditions for MUAP analysis. The recordings were made at low (just above threshold) voluntary and constant level of contraction. A standard concentric needle electrode was used. The EMG signals were recorded from bicep brachii of bicep. The high and low pass filters of the EMG amplifier were set at 2 Hz and 10 kHz for 11 ms duration with sampling frequency of 10 kHz [8].

2.3 Proposed Method

In this paper, we describe each of the following steps involved for the classification of the acquired EMG signal. These steps are: (A) Base line shift for accurate estimation of parameters (B) Wavelet analysis for EMG signal (C) Various parameters/features calculation (D) Classifier Design. Figure 2 describes the process.

2.3.1 Base Line Shifting

Generally human being produces static current which will interfere while recording EMG signal. Due to this problem EMG signal shifts upper side from the base line. This defect occurs due to the muscle tension, body movement or some environmental noise.

2.3.2 Wavelet Analysis for EMG Signal

Wavelet analysis was performed for short time period and the size of window is can be varied by using the real time frequency analysis. So the result at low and high frequency will gives us the same result. Here we have used the Symlet mother wavelet of order five for EMG signal analysis. The level of decomposition is three. Each wavelet is based on function called mother wavelet and subset are operated with scaling (a) and translation (b) is represented by equation

$$\emptyset(t) = \frac{1}{\sqrt{a}}\emptyset\left[\frac{t-b}{a}\right] = \emptyset(scale, position, t)$$

To extract the required EMG band, wavelet decomposition [11–13] is done which has certain advantages over other analysis like time–frequency localization, multi rate filtering and scale-space analysis. Variable window size is taken to compress and stretches the wave with decomposition is done up to third level to get the required bands by means of multi resolution analysis using complementary low and high pass filter. The original signal is decomposed and two sequences will be found out with high and low resolution frequency. Then the lower frequency component is decomposed to get the second level of component of higher and lower frequency and the same process will carry on up to fourth level to get the required frequency levels. Pictorial representation of wavelet decomposition is presented in Figs. 3 & 4. After the step the statistical parameters will be calculated for the bands as described below. The required band for our analysis after reconstruction is $S = Ca_3 + Cd_1 + Cd_2 + Cd_3$.

where Ca is the approximated value and Cd are the detail values (Figs. 2, 3, 4).

2.3.3 Feature Extraction and Calculation

In this paper, we are calculating different entropies for classifying above referred EMG signals. Entropies are stated below:

Lyapunov Exponent

It is measure of rate of separation of infinitesimally close trajectories for dynamical system. It gives the rate of exponential divergence from perturbed

Fig. 2 Block diagram describing the process

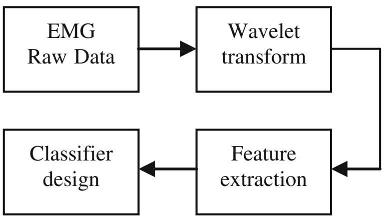
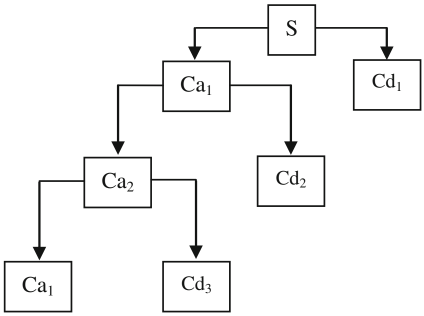


Fig. 3 Wavelet decomposition up to level three



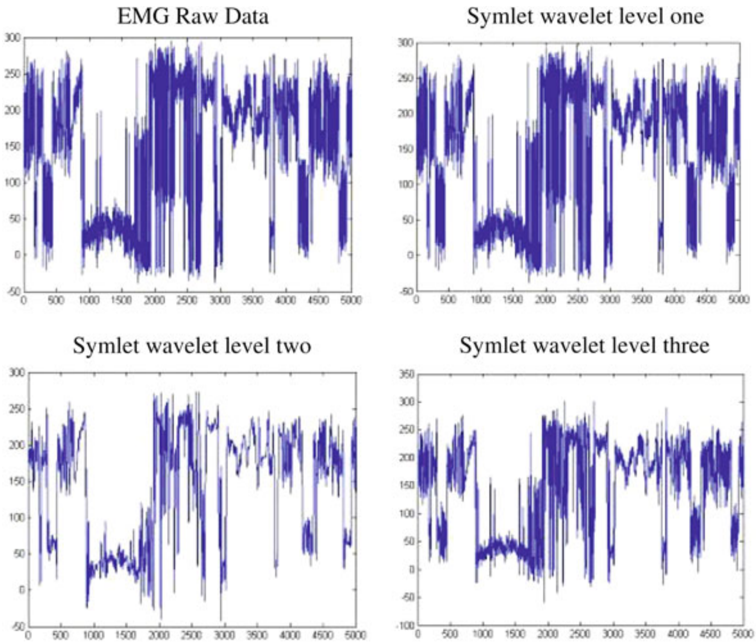


Fig. 4 Various levels of wavelet analysis

initial conditions. It is a measure of the rate at which nearby trajectories in phase space diverge. Thus a value of +1 means separation of nearby orbits doubles on the average in the time interval between data samples.

$$\lambda = \lim_{\substack{t\to\infty \\ |\Delta x_0|\to 0}} \frac{1}{t} ln \frac{|\Delta x(X_0, t)|}{|\Delta x_0|}$$

The displayed error is estimated from 2.5 times the standard deviation of the slope divided by the square root of the number of trajectories followed. It is given by following [6].

Capacity Dimension

It is also known as Hausdorff dimension. It is calculated successively dividing the phase space with embedding dimension D into equal hyper cubes that are occupied with data points with respect to the log of the normalized linear dimension of hypercubes. The capacity dimension of space X is a real number $d_{capacity}$ such that if n(∈) denotes the minimum number of open sets of diameter less than or equal to ∈, then n(∈) is proportional to $\in^{-D}$ as $\in \to 0$. It can be given as below.

$$d_{capacity} = -\lim_{\in \to 0+} \frac{\ln N}{\ln \in}$$

Where N is the number of elements forming a finite Space and ∈ is a bound on the diameter of the sets involved (informally, ∈ is the size of each element used to cover the set, which is taken to approach 0) if the limit exists.

Correlation Dimension

It is a method to determine the dimension of space occupied by sets of random points [9, 10]. It provides a good measure of the complexity of the dynamics, i.e. of the number of active degrees of freedom. The correlation dimension is a characteristic of the underlying invariant measure μ. In a certain sense it characterizes how smoothly μ is distributed over the attractor: if μ is a point measure, then $\alpha = 0$, and if μ is absolutely continuous with respect to Lebesgue measure, then α equals the topological dimension d of X. These are two boundary cases, in general $0 < \alpha < d$. This paper explains the correlation dimension in following way

$$\mathrm{C(r)} = \int \int_{\Omega\times\Omega}^{\infty} I(\|y_1 - y_2\| \leq r) d\mu(y_1) d\mu(y_2)$$

where I denote an indicator function.

Correlation Matrix

Singular Value decomposition begins with the correlation matrix which is a two dimensional matrix formed by putting the value of the correlation function with tau = 0 along the diagonal and putting of the value of the correlation function with tau = 1 to the right and left filled. The Eigen values are determined.

Probability Distribution Function

It shows the probability of occurrence of certain data values. We can specify a number of bins between 2 and 128 into which data points are sorted according to their value. The bins all have the same width and are spread uniformly between the lowest and highest data value.

$$\mathrm{F(x)} \ = \ \mathrm{F(x)} = \int_{-\infty}^{\chi} f(t)dt$$

Hurst Exponent

It is referred to as the "index of dependence", and is the relative tendency of a time series to either strongly regress to the mean or 'cluster' in a direction. The Hurst exponent is used as a measure of the long term memory of time series, i.e. the autocorrelation of the time series.

$$E\left[\frac{R(n)}{S(n)}\right] = Cn^{H} asn \rightarrow \infty$$

where

$\left[\frac{R(n)}{S(n)}\right]$ is the rescaled range.

E[x] is the expected value.

n is the time of the last observation

(e.g. it corresponds to X_n in the input time series).

C is a constant.

2.3.4 Support Vector Machine as a classifier

In this paper, after feature extraction, feature matrix is normalized in order to have features in a same range. Simply, SVM classifier [14, 15] was trained by the train–train data and then used for classifying the train-test data. SVM creates a separating hyper plane between two clusters of data, thus classifying data falling into two different classes. In non-linear classification, same operation was performed by non-linear kernel function, in our case, Gaussian-RBF kernel, given by

$$k\left(x_i, x_j\right) = \exp\left(-\left\|x_i - x_j\right\|^2 \Big/ 2\sigma^2\right)$$

In this work we take two different group i.e. normal and abnormal (Myopathy and Neuropathy) signal for the classification purpose. We take 70 % of signal to train the expert model using SVMStruct a structure with the Kernel Function fields

[14, 15]. Kernel function Value is Gaussian Radial Basis Function kernel (‘rbf’) function handle with a scaling factor, sigma which is varying from 0 values to 10 and takes those which give maximum accuracy.

3 Result and Discussion

We are classifying EMG signals for differentiating different diseases from the normal. Our classification uses features which has described above. So we are spotting some of the features which produce demarcation between different subjects. Then we train and test the SVM classifier with various extracted features like for EMG. The classifier was also cross validated by taking feature vector in different combinations

From the result we state that Lyapunov Exponent and Hurst Exponent is the best parameter to classify the EMG signals when taken individually and the classification accuracy is around 94.76 %. By taking the combined features Hurst Exponent and Lyapunov Exponent gives us classification accuracy around 86.48 %.

4 Conclusion

In this present study, we address the problem taking the statistical nonlinear properties after wavelet reconstruction. The parameters like LLE, CD, CAD, HE, CM will gives us a clear distinction for classifying and detecting the abnormality in EMG signals. The technique described is a cost effective as well as time effective approach to design an automated classifier for Muscular Atrophy Diseases Diagnosis Approach. The future work lies with the classifier i.e. we will take different classifier and compare the results with each other.

References

1. Boisset S, Goubel F (1972) Integrated electromygraphy activity and muscle work. J Applied Physiol 35:695–702
2. Plonsey R (1974) The active fiber in a volume conductor. IEEE Trans Biomed Eng 21:371–381
3. De Luca C, (2006) ‘Electromyography’. Encyclopedia Of Medical Devices and Instrumentation. In: John G (ed) Webster, Wiley, New york
4. DeLuca CJ (1993) Use of the surface EMG signal for performance evaluation of back muscle. Muscle Nerve 16:210–216
5. Jones RS et al (1988) On-line analysis of neuromuscular function. In: Proceeding of IEEE Engineering in Medicine Biology, vol 10, p 1607

6. Mengi Y, Liu B, Liu Y (2001) A comprehensive nonlinear analysis of electromyogram engineering in medicine and biology society. In: Proceedings of the 23rd annual international conference of the IEEE, vol 2. pp 1078–1081 (Nov 2002)
7. Small GJ, Jones NB, Fothergill JC, Mocroft AP (2002) Chaos as a possible model of electromyographic activity simulation 98. International conference on (Conf. Publ No 457) pp 27–34
8. Erfanian A, Chizeck HJ, Hashemi RM (1996) Chaotic activity during electrical stimulation of paralyzed muscle engineering in medicine and biology society bridging disciplines for biomedicine. In: Proceedings of the 18th Annual International Conference of the IEEE, vol 4. pp 1756–1757 (Aug 2002)
9. Bodruzzaman M, Devgan S, Kari S (1992) Chaotic Classification Of Electromyographic (Emg) Signals Via Correlation Dimension Measurement Southeastcon '92. In: Proceedings IEEE, vol 1. pp 95–98 (Aug 2002)
10. Yang H, Wang D, Wang J (2006) Linear and non-linear features of surface EMG during fatigue and recovery period. Engineering in Medicine and Biology Society, IEEE-EMBS 2005.In: 27th annual international conference, pp 5804–5807, Apr 2006
11. Englehart K, Hudgins B, Parker PA (2001) A wavelet-based continuous classification scheme for multifunction myoelectric control. IEEE Trans Biomed Eng 48(3):302–311
12. Aldroubi A, Unser M (eds) (1996) Wavelets in medicine and biology. CRC Press, New York
13. Rioul O, Vetterli M (1991) Wavelets and signal processing. IEEE Signal Process Mag, pp 14–38 Oct 1991
14. DeLuca CJ (1993) Use of the surface EMG signal for performance evaluation of back muscle. Muscle Nerve 16:210–216
15. Jones RS et al (1988) On-line Analysis of Neuromuscular Function. In: Proceeding IEEE engineering in medicinal biology, vol l0. p 1607 (Nov 1988)

LOG-Grid Based Off-Line Signature Verification System

B. H. Shekar and R. K. Bharathi

Abstract This paper presents a LOG-processed local feature based off-line signature verification system. The proposed approach has three major phases: Preprocessing, Feature extraction and Classification. The contour of the signature is divided into equal grids of size M $\times$ M and the 4-directional chain code histogram of each grid is extracted, thus fed into laplacian of Gaussian filter to enhance the feature representing the signature image. These enhanced feature vectors are fed as an input to the multi-layer-perceptrons (MLP) for training purpose. The MLP is used as a recognition tool and trained with different number of training samples including genuine, skilled and random forgeries and hence tested. This model was successfully tested on two datasets, namely, CEDAR, a publicly available English signature dataset and MUKOS, a regional language (Kannada) dataset and compared with well-known approaches to exhibit the performance of the proposed approach.

Keywords Laplacian of Gaussian · Chain code · Histogram · MLP · Off-line signature

B. H. Shekar (✉)
Department of Computer Science, Mangalore University, Mysore, Karnataka, India
e-mail: bhshekar@gmail.com

R. K. Bharathi
Department of Master of Computer Applications, S J College of Engineering, Mysore, Karnataka, India
e-mail: rkbharathi@hotmail.com

Mohan S. and S. S. Kumar (eds.), *Proceedings of the Fourth International Conference on Signal and Image Processing 2012 (ICSIP 2012)*, Lecture Notes in Electrical Engineering 222,
DOI: 10.1007/978-81-322-1000-9_31, © Springer India 2013

1 Introduction

Biometric recognition forms a strong link between a person and his/her identity as the biometric traits which cannot be easily shared, lost or duplicated. Handwritten signature happens to be the most natural behavioural biometric which generally established mode of providing the personal identity for authentication. Analysis of the signature can only be applied when the person is conscious and capable to sign in the regular pattern. However for many reasons the signature verification cannot be easily considered as a trivial pattern recognition problem [7, 8]. The major threat/challenge in Signature verification and recognition is the Forgery of the Signature, an illegal modification or reproduction of a signature. A signature is also considered forged if it is claimed that it was made by someone who did not make it [17]. Here, forgeries can be classified into three basic types: Random forgery: when the forger nor has the access to the genuine signature neither has any information about the authors name but reproduces a random signature, Simple forgery: where the forger has no access to the sample signature but knows the name of the signer, and lastly, Skilled forgery: when forger reproduces the signature, having the access to the sample genuine signature [1, 16].

Based on the procedure of acquiring, the signature is classified as On-line signature and Off-line signature. On-line signature acquired at the time of its registration provides the intrinsic dynamic details viz: *velocity, acceleration, direction of pen movement, pressure applied and forces*, where as in Off-line signature, the static image of the signature is captured once the signing process is complete. In spite of the fact that Off-line signature lacks the dynamic details, it withholds many of the global and local features viz: *signature image area, height, width, zonal information, characteristics points such as end points, cross points, cusps, loops, presence and absence of zonal information, so on*. The off-line signature verification includes learning the pattern of the signature image and there by to recognize the test image is genuine or forge, based on the features extracted and trained. We have proposed an efficient technique called *LOG-MLP* model for verification and recognition of the signatures, with high performance rate. The details of the proposed model are brought out in the following sections. The paper is organized as follows: The Review of the related works are brought down in Sect. 2. In Sect. 3, the proposed model is explained. In Sect. 4, the experimental set-up along with a brief note on the dataset used in the experimentation is given and the experimental results are given followed by conclusion in Sect. 5.

2 Review of Related Work

Off-line signatures are the static 2D-image of the registered signature at a certain point of time. Processing the signatures imply, accessing the images in the absence of dynamic information, which increases the complexity of verifying for genuine or classifying them from the forge. Hence, quality enhancement of the image, preprocessing and few morphological operations on the signature image plays the vital role in designing the efficient technique of feature extraction for off-line signature verification. Some of the well accepted off-line signature verification models based on varying features and feature selection techniques is reviewed here. Nguyen and Blumenstein [14] derived the features from the total energy a writer uses to create their signature following the projections (both horizontal and vertical), which focuses on the proportion of the distance between the key strokes. They also described a grid-based feature extraction technique that utilizes directional information extracted from the signature contour and by applying 2D Gaussian filter on the feature matrices. Kalera et al. [11] extracts the features on a quasi-multi-resolution technique using GSC (Gradient, Structural and Concavity) which was earlier used for identifying the individuality of handwriting, mapping from the handwriting domain to the signature domain on CEDAR dataset. Chen and Srihari [3] extracted the contours of the signature and combine the upper and lower contours neglecting the smaller middle portion to define a pseudo-writing-path. To match two signatures, DTW (Dynamic Time Warping), a non-linear normalization is applied by segmenting them into curves, followed by extraction of features using Zernike moments (shape descriptor). Kumar et al. [12] comes with the objective of the work in two folds, one, to propose a feature set based on signature morphology and the second, to use a feature analysis technique to get a more compact set of features. Experiments were conducted on CEDAR dataset to exhibit the performance of the system. Ismail et al. [9] present a novel off-line signature identification based on chain codes and fourier descriptors. Here they classify the signature image identification into two different problems: Recognition and Verification. For recognition they have used Principal Component Analysis and Multilayer feed forward artificial neural network for verification. Ghandali et al. [5] extracts features by allowing the Discrete Wavelet Transform (DWT) to access details of signatures, then several registered instances of each person signature are fused together to generate reference pattern of person signatures. An euclidean distance measure is employed to verify the genuine signature and worked on Persian signatures. Marianela et al. [13] concentrates on the rotated signature images and hence a technique for rotation invariant feature based on discrete fourier transform is designed. Jena et al. [10] proposed to extract the features, by simple horizontal and vertical splits of signature sample. From each split, the geometrical centers points are considered as features, with thirty horizontal and thirty vertical splits. Gilperez et al. [6] works on the local image level and encodes directional properties of signature contours and the length of regions enclosed inside the signatures. Here the signature is considered to be a shape like texture that is described with probability distribution functions (PDFs). A new approach based on subset of the line, concave and convex family of curvature features is used to represent the signature, resulting in two step transitional features introduced by Zois et al. [21]. Similar to this, Tselios et al. [20] considered relative pixel distribution over predetermined two and three-step paths along the signature trace that result in a model for estimating the transitional probabilities of the signature stroke, arcs and angels as features. Shekar and Bharathi [18] concentrated on reducing the dimension of feature vectors, preserving the effective features obtained through eigen-signatures and later extended to Kernel eigen-signatures based on Kernel-PCA [19]. Once the features are extracted and reduced with the dimension, better choice of classifier is of most importance too. Generally, the researchers work with a single classifier: namely, Support Vector machine, Neural Netwrok Based, Distance based and so on, suitable for the features extracted from the signature image for verification and identification. Kisku et al. [4] reported a weighted fusion of multiple classifiers using geometric and orientation features and tested on IIT Kanpur off-line signature database. Batista et al. [2] proposed the Hybrid generative-discriminative ensembles of classifiers for off-line signature verification, where the classifier selection process is done dynamically. Thus feature extraction and representation plays a vital role in off-line signature verification and recognition. In this context, we have introduced a new LOG-Grid based approach for signature representation followed by MLP for classification. The details are presented below.

3 Proposed Model

The Proposed LOG-MLP model of off-line signature verification involves three major phases: preprocessing and contour extraction, creation of 4-directional chain code histogram on the grid of the signature contour, feature extraction after applying the Laplacian of Gaussian filter. The feature matrix are used to train the classifier multilayer perceptrons (MLP-NNFF) and hence perform the verification and recognition of the off-line signatures. In the preprocessing stage, the signature images are binarised to remove the complex background and later the contour of the signature image is drawn, which happens to be the input image for further processing.

3.1 LOG-Grid Based Feature Extraction

The signature image contour (as shown in Fig. 1) is the input to the LOG-Grid based feature extraction technique and the algorithm is as follows:

(a) Divide the input signature contour image into 12×12 grid blocks.
(b) The 4-directional chain-code histogram is generated by tracing the contour along each block as shown in Fig. 2. Tracing in 4-directions results in four matrices of size 12×12 created for each of the directions. Let us name these matrices as $H_m, V_m, L_m\, and\, R_m$ created by horizontal, vertical, left and right directional tracing respectively.

(c) Apply the Laplacian of Gaussian filter on each of the directional 12 × 12 matrices obtained in step (b) with $\sigma = 0.5$ and the mask of size 5 × 5.

$$A^*_{i,j} = \sum_{di=-\infty}^{+\infty} \sum_{dj=-\infty}^{+\infty} A_{i+di,j+dj} \frac{1}{2\pi\sigma^2} e^{-\frac{d_i^2+d_j^2}{2\sigma^2}} \tag{1}$$

(d) To normalize the elements in the matrix, each element value of each matrix obtained by the step (c) is divided its value by the maximum value of the four matrices.

$$A_{i,j} = \frac{A_{i,j}}{max(H_m, V_m, L_m, R_m)} \tag{2}$$

(e) Two additional matrices S_m and P_m are calculated using the existing 4 directional chain code matrices by pairing the H_m with V_m and L_m with R_m as the proportions between perpendicular directions are relatively stable features and leads to better verification accuracies [15]:

$$S_m = \begin{cases} \frac{max(H_m, V_m)}{min(H_m, V_m)} & if\ max(H_m, V_m) \neq 0 \\ 0 & otherwise \end{cases} \tag{3}$$

$$P_m = \begin{cases} \frac{min(L_m, R_m)}{max(L_m, R_m)} & if\ max(L_m, R_m) \neq 0 \\ 0 & otherwise \end{cases} \tag{4}$$

Thus, by appending all the above six matrices ($H_m V_m L_m R_m S_m P_m$) the feature vector of size 12 ×12 × 6 is created. Initially the experiments with 8 × 8, 10 × 10 grid configurations were conducted, and later extended to 12 × 12 grid size, which produced the acceptable results and hence continued with ÏČ = 0.5 and LOG mask 5 × 5. The feature repository consists of all the signer's genuine signature feature set followed by their skilled forgery features.

Fig. 1 **a** Initial image. **b** Binarised and noise removed of image (**a**). **c** Contour of image (**b**)

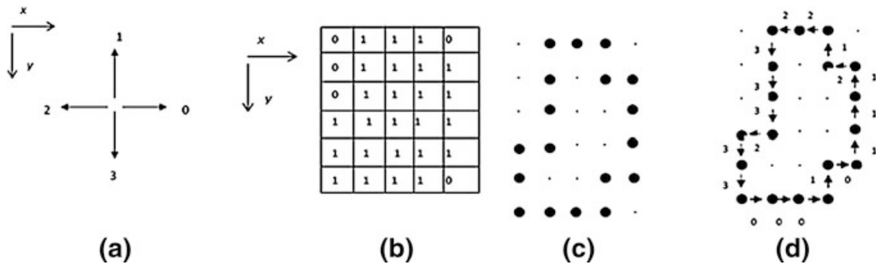


Fig. 2 **a** Direction number for 4-directional chain codes. **b** A sample 4-connected object. **c** Boundary of (**b**). **d** Obtaining the 4-connected chain code for the object in (**b**) and (**c**)

4 Experimental Settings and Result Analysis

Automatic signature verification can produce two types of errors: Type-1 errors which concerns on the false rejection of the genuine signatures to be forge and can be termed as FRR (False Rejection Rate); Type-2 errors concerned on the false acceptance of the forged signature as genuine and is termed as FAR (False Acceptance Rate) [8]. Hence, FAR is the percentage ratio of the total number of accepted forgeries to the total number of tested forgeries, whereas FRR is defined by the percentile ratio of total no. of genuine rejected to the total number of tested genuine. The performance of signature verification system is estimated in terms of FRR and FAR and good trade-off need to be made between them as any reduction of FAR increases FRR and vice-versa. Selection of optimal parameters for MLP is always a challenging task as it is solely application (feature value) based and can be decided only after trial experimentations. Here we have employed 15 hidden layers, with Sum of Squared Error (SSE) as an error performance function with a goal of 0.001 and a maximum of $4,500$ epochs. The training is carried out by creating feed-forward back propagation network and later a simulator is used to test the data along with the network created while training. We have conducted the experimentations on benchmark signature dataset to reveal the performance of the proposed model with varying number of training and testing signature combinations, consisting of genuine signatures with random forgery and skilled forgery. To demonstrate the results and the performances of the proposed model, we have considered publicly available dataset and our own regional language signature dataset. Experiments were conducted with varying number and combinations of samples for training and hence for testing. The dataset includes, CEDAR [3] and MUKOS [18] with a total of *2640* signatures in *CEDAR* and *2110* signatures in *MUKOS* as listed in Table 1.

Experimentation on CEDAR Database: Centre of Excellence for Document Analysis and Recognition (CEDAR), at SUNY at Buffalo, has built the off-line signature corpus with 55 signers, each with 24 genuine signature samples. To obtain forgeries (skilled), some 20 arbitrary signers were asked to skill fully forge the signature in the corpus, each with 24 samples, resulting the CEDAR signature corpus, a total of 2640 samples. Sample genuine and skilled forgery signatures are

Table 1 Number of samples in the datasets used in the experimentation

Dataset	Number of signers	Number of genuine signatures	Number of skilled forgery	Total number of signatures
CEDAR	55	24	24	2640
MUKOS-Set-1	30	30	15	1350
MUKOS-Set-2	19	20	20	760

as shown in Fig. 3. The CEDAR Signature database is available on http://www.cedar.buffalo.edu/NIJ/publications.html/.

Initially we started with first 10 genuine and first 10 skilled forgery to train the network and test with the remaining 14 samples of genuine and forgery. Later we extended it train with 16 + 16 samples of genuine and forge and tested with 8 genuine + 8 skilled forged samples for all the 55 signers in the dataset. We also experimented by training the network with random forgery (genuine of other signers in the corpus) along with genuine samples, where we tried with different combinations as the random forgery was selected randomly. At this point we trained the network with 16 genuine samples and 200 random forgery (randomly selecting 4 genuine samples of other 50 randomly chosen signer) and tested with the remaining 8 genuine and 16 random forgery (4 randomly selected samples of remaining 4 singers other than the trained and targeted signer). Here to justify the random selection, experimentation with 8 different combinations of randomly selecting the forge was carried out. The best performance is resulted below in the Table 2, which reveals both the comparative analysis of the results in the literature to that of the proposed model on CEDAR dataset.

Experimentation on MUKOS database: MUKOS database consists of two sets, with Set-1 consisting of 1350 signatures from 30 signers and Set-2 760

Table 2 Experimental results obtained for CEDAR dataset:

Proposed by	Feature type	Classifier	Accuracy	FAR	FRR
Kalera et al. [11]	Word shape (GSC)	PDF	78.50	19.50	22.45
Chen and Shrihari [3]	Zernike moments	DTW	83.60	16.30	16.60
Kumar et al. [12]	Signature morphology	SVM	88.41	11.59	11.59
Proposed model	**LOG-Grid**	**MLP**	**91.72**	**7.28**	**7.68**

Fig. 3 Sample genuine signatures and corresponding skilled forgery from CEDAR corpus

signatures. In Set-1 we collected 30 genuine signatures and 15 skilled forgeries from each signer. Each genuine signature was collected using black ink on A4 size white paper featuring 14 boxes on each paper. Once the genuine signatures were collected by all thirty signers, the forgeries were produced imitating a genuine signature from the static image of the genuine after a time gap where they were allowed to practice the forgery of other signers (other than genuine signers). In Set-2 we collected 20 genuine and 20 skilled forgery samples from 38 (19 + 19) different individuals. Skilled forgery was obtained by arbitrarily selecting the people who in-practice uses English for their genuine signature. These signatures were acquired with a standard scanner with 75 dpi resolution in an 8-bit gray scale image. Figure 4 gives the sample kannada genuine signature and theircorresponding skilled forgery. Even with the MUKOS dataset, we started with first 10 genuine and first 10 skilled forgery to train the network and test with the remaining 20 of set-1 and 10 in set-2 samples of genuine and forgery. Later we performed training with 15 of set-1 and 12 of set-2 samples of genuine with 10 skilled forge from each set correspondingly, thereby testing with remaining 15 genuine of set-1 + 5 skilled forged samples and 8 genuine of set-2 with 10 skilled forge for all the 30 + 19 signers in the dataset. Experimentation was carried out considering training random forgery, the genuine sample of other signer along with genuine samples, randomly selecting the training and testing samples. We have trained the network with 15 of set-1 and 12 of set-2 genuine samples and 200 random forgery (randomly selecting 4 genuine samples of other 25 randomly chosen signer) of set-1 and tested with the remaining 15 of set-1 and 8 of set-2 genuine and 16 random forgery (4 randomly selected samples of remaining 4 singers other than the trained and targeted signer). Here to justify the random selection, experimentation with 5 different combinations by randomly selecting the forge was carried out and the best performance is shown in table. With the LOG-Grid based feature being trained on MLP, has improved the error rates as the adequate samples are trained. It is observed that efficient feature extraction and recognition system has reduced the FAR from 11.07 to 6.59, there by reduction in the rate of forge being accepted as genuine. The performance of the proposed model on MUKOS is tabulated in Table 3. Experimental results of the proposed LOG-Grid feature based model trained and tested with random forgery and their best performances of the experimentations conducted for 5 times by randomly selecting the random forgery is listed in Table 4 [Best result is listed].

Table 3 Experimental results for MUKOS dataset

Proposed by	Feature type	Classifier	Accuracy	FAR	FRR
Shekar and Bharathi [18]	Shape based eigen signature	Euclidean distance	93.00	11.07	6.40
Proposed model on Set-1	**LOG-Grid based**	**MLP**	**93.41**	**6.59**	**5.59**
Set-2			**92.66**	**7.35**	**6.78**

Table 4 Best performance of the proposed model [trained and tested with **Genuine and Random Forgery**]

Dataset	Accuracy	FAR	FRR
CEDAR	93.46	1.04	6.60
MUKOS-Set-1	95.21	2.12	4.88
-Set-2	95.19	1.94	4.82

Fig. 4 Sample genuine signatures and corresponding skilled forgery from MUKOS (Kannada) signature dataset

5 Conclusion

In this work, we developed LOG-Grid based feature extraction method followed by the verification of signatures using MLP. The proposed model's performance is demonstrated on publicly available English dataset: CEDAR and a local language, Kannada signature dataset (MUKOS). Even though the model is highly computational as we are dealing with 4-directional chain code on each dominant pixel, the performance is better than any other methods. Experimental results on the standard dataset reveals the performance of the system.

References

1. Arya MS, Inamdar VS (2010) Article: a preliminary study on various off-line hand written signature verification approaches. Int J Comput Appl 1(9):55–60
2. Batista L, Granger E, Sabourin R (2012) Dynamic selection of generative-discriminative ensembles for off-line signature verification. Pattern Recogn 45:1326–1340
3. Chen S, Srihari S (2005) Use of exterior contours and shape features in off-line signature verification. In: ICDAR, pp 1280–1284
4. Kisku DR, Ajita Rattani PGJKS (2011) Offline signature verification using geometric and orientation features with multiple experts fusion. IEEE International Conference on Network and Computer Science (ICNCS/ICECT 2011) 5, 269–272 (October 2011)
5. Ghandali S, Moghaddam M (2009) Off-line persian signature identification and verification based on image registration and fusion. J Multimedia 4(3):137–144
6. Gilperez A, Alonso-Fernandez F, Pecharroman S, Fierrez J, Ortega-Garcia J (2008) Off-line signature verification using contour features. In: Proceedings of ICFHR
7. Hou W, Ye X, Wang K (2004) A survey of off-line signature verification. In: Proceedings of international conference on intelligent mechatronics and automation, pp 536–541
8. Impedovo D, Pirlo G (2008) Automatic signature verification: the state of the art. IEEE Trans Syst Man Cybern Part C Appl Rev 38(5):609–635
9. Ismail I, Ramadan M, El-Danaf T, Samak A (2010) An efficient off line signature identification method based on fourier descriptor and chain codes. IJCSNS 10(5):29

10. Jena D, Majhi B, Panigrahy S, Jena S (2008) Improved offline signature verification scheme using feature point extraction method. In: 7th IEEE international conference on cognitive informatics (ICCI 2008), pp 475–480
11. Kalera MK, Srihari S, Xu A (2003) Off-line signature verification and identification using distance statistics. Int J Pattern Recognit Artif Intell 18(7):228–232
12. Kumar R, Kundu L, Chanda B, Sharma JD (2010) A writer-independent off-line signature verification system based on signature morphology. In: Proceedings of the first international conference on intelligent interactive technologies and multimedia (IITM '10), ACM, New York, pp 261–265
13. Marianela P, Juan C, Belaïd, A et al (2011) A circular grid-based rotation invariant feature extraction approach for off-line signature verification. In: ICDAR, pp 1289–1293
14. Nguyen V, Blumenstein M (2011) An application of the 2d gaussian filter for enhancing feature extraction in off-line signature verification. In: ICDAR'11, pp 339–343
15. Nguyen V, Blumenstein M, Leedham G (2009) Global features for the off-line signature verification problem. In: IEEE Computer Society ICDAR'09, pp 1300–1304
16. Sachdeva S et al (2011) An offline signature verification technique. Webmedmentral 2(5):1–10
17. Saikia H, Sarma KC (2012) Article: approaches and issues in offline signature verification system. Int J Comput Appl 42(16):45–52
18. Shekar BH, Bharathi RK (2011) Eigen-signature: a robust and an efficient offline signature verification algorithm. In: International conference on recent trends in information technology (ICRTIT), pp 134–138
19. Shekar BH, Bharathi RK, Sharmilakumari M (2011) Kernel eigen-signature: an offline signature verification technique based on kernel principal component analysis. In: Emerging applications of computer vision (EACV-2011), bilateral Russian-Indian scientific workshop (2011), pp 37–44
20. Tselios K, Zois EN, Nassiopoulos A, Economou G (2011) Fusion of directional transitional features for off-line signature verification. In: International joint conference on Biometrics, pp 1–6
21. Zois E, Tselios K, Siores E, Nassiopoulos A, Economou G (2011) Off-line signature verification using two step transitional features. In: Proceedings of the 12th IAPR conference on machine vision applications, Nara, Japan, pp 295–298

Colorogram: A Color Feature Descriptor for Human Blob Labeling

Vejey Subash Gandyer

Abstract Tracking is one of the most important research fields for computer vision. Human Identity tracking is significant in the field of surveillance where a foolproof system is required that not only finds an existence of some person behaving suspicious but also his identification. In this paper, a novel feature descriptor by the name Colorogram has been proposed for labeling human blobs in videos. Two constraints were introduced into the system by way of color information and spatial spread of pixels in a new coordinate system. This feature descriptor helped in identifying/labeling human blobs with a priori knowledge of the scene. Two color spaces (HSV, RGB) and two histogram distance measures (Intersection, Euclidean) were considered for computation. The proposed Colorograms were compared with the conventional Histograms and Spatiograms. Experiments were conducted and the results have been tabulated. It was inferred that the new feature descriptor had classified humans with an acceptable precision and recall rates.

Keywords Human identity · Colorogram · Tracking · Feature extraction · Histogram · Blob · Spatiogram

1 Introduction

In recent years, automated surveillance has gained much attention. Continuous monitoring of a place of interest is needed and this involves deploying a human to watch videos all day long, which is really a daunting task. So there is a need to

V. S. Gandyer (✉)
Department of Computer Science and Engineering, KCG College of Technology, Anna University, Chennai, India
e-mail: subashgandyer@ieee.org

Mohan S. and S. S. Kumar (eds.), *Proceedings of the Fourth International Conference on Signal and Image Processing 2012 (ICSIP 2012)*, Lecture Notes in Electrical Engineering 222,
DOI: 10.1007/978-81-322-1000-9_32,

assist him in monitoring people. Although it is much difficult to fully automate without human intervention, certain simple questions like "How many people are there in the scene?", "What is the identity of each person?", "Were this person been here before?", "If so when is the last time he was here?" and "With whom has he interacted?" can be answered to assist security personnel. An attempt to develop a system for people monitoring in public places has been tried with a view to answer these vital questions. A simple framework for surveillance has been proposed with various functionalities like tracking multiple objects in the scene, keeping track of objects by labeling them even under occlusion, identifying humans even they appear in the scene after sometime. First the foreground is extracted using the background subtraction methods discussed in [1–3]. The blobs are extracted from the background and only the blobs that are above an experimentally tested threshold, are considered for further processing. Tracking of objects is achieved by using the colorogram model where each person is visually interpreted as a color histogram model based on the pixel color of the blob and spatial spread of the pixels. This is stored as a model and the blobs coming in from successive frames are checked for similarity. If there is a high match, the blob is assigned to the corresponding model and if that blob does not match with any models available then it is stored as a separate model. This process is repeated for all frames in the live feed. Thus, tracking of objects in the scene i.e., maintaining correspondence between blobs of two frames is achieved using colorograms.

2 Related Work

One important issue when modeling the visual appearance of people is the complexity of the model. The underlying assumption is that the appearance of a moving object varies in a regular manner over time and can therefore be modeled and predicted from frame to frame. There have been a quite a number of contributions in the tracking domain. For the purpose of tracking, several papers have been explored and these are one of the few to mention. Background is modeled in a non-parametric way in [1, 2] for foreground separation. Capedallas and DeMenthon labeled humans in an indoor environment in [4]. Bremond and Thonnat [5] proposed a system for preventing vandalism in metro stations. Haritaoglu et al. [6] studied the real-time surveillance of people. Desah and Salih [3] extracted objects in real-time using background subtraction. Various data structures such as Graphs and Trees were employed in studying image segmentation and analysis. Of trees, Binary trees were used widely. Color feature descriptors based on Quad trees were proposed in [7] and [8]. Spatiograms, another way of representing color information with geometry information preserved in the descriptor were introduced by Birchfield and Rangarajan in [9].

3 Colorogram

3.1 Colorogram Construction

Colorogram: The colorogram h of image I denoted as $h_I(c_i)$ is defined such that $h_I(c_i)$ gives for any pixel in I, the probability that the color of the pixel is c_i. It is defined as the function of the color information and polar information. $C(H_I)$ is the histogram of the image I that gives the summation of the number of colors present in the image.

$$h_I(c_I) = \omega_1 c(H_I) + \omega_2 c(\rho_{max}, \theta) \tag{1}$$

$$c(H_I) = \sum_i \frac{H_i(c_i)}{|I|} \tag{2}$$

ω_1 is the weighted factor for the introduced Color constraint and ω_2 is the weight for Polar coordinate geometry. It refers to the distribution of the colors in the blob i.e., how many times does a color appear in a scene. After the motion detector detects the motion and extracts the foreground and turns into a blob, the frequency of color distribution of the blob i.e., its colorogram is computed and stored as a model for further analysis.

Colorogram is the color distribution of the spatial points of an image in the polar coordinates. Once the blob is extracted from the background subtraction method, the contour of the blob is considered for constructing the colorogram model as shown in Fig. 1. This model works on the assumption that all blobs should be a closed contour. Open contour models need to be closed by approximation or interpolation algorithms before constructing the colorogram for that blob.

Algorithm 1: Colorogram Construction

Input : Blob using adaptive background subtraction
Output : Colorogram for the blob

1. Sample N points inside closed contour of blob.
2. Calculate centroid of the blob & set it as origin.
3. Translate contour points into polar coordinates.
4. Draw a circle with centroid as centre circumscribing the blob.
5. Divide the circle into bins taking into account number of colors to be chosen.
6. Store number of points located in each partition into its corresponding bin.
7. Output Colorogram for the blob

Fig. 1 Algorithm for construction of colorogram

Two constraints are introduced into this environment to create this colorogram. The first constraint is the spatial constraint that determines the spatial spread of the pixels in the image. Color information is the second constraint that is introduced into this system. The introduction of first constraint is justified below. N points are sampled in the closed contour and the resultant sampled points on the blob can be represented as in (3). The centroid of the blob is computed. As this colorogram model is based on the Polar space coordinate system, the centroid of the blob (x_c, y_c) is set as the origin reference for the new coordinate system. Once the centroid is fixed as centre at (0, 0), the sampled points P is translated into Polar coordinates as in (4)–(6). A circle is drawn circumscribing the blob with (x_c, y_c) as centre and ρ_{max} as radius where $\rho_{max} = \max \{\rho_i\}$, i = 1,2,…,N. This circle is partitioned into 'a' × 'b' bins with 'a' number of bins as a function of (ρ_{max}, θ) and 'b' number of bins as a function of C_{index} that will be explained below. Once the bins are constructed in the above said manner, they are filled with respect to the above said factors and a colorogram distribution is constructed.

$$P = \{(x_1, y_1), \ldots, (x_N, y_N)\},\ (x_i, y_i) \in R^2 \tag{3}$$

$$P = \{(r_1, \rho_1), \ldots, (r_N, \rho_N)\},\ (r_i, \rho_i) \in R^2 \tag{4}$$

where,

$$r_i = \sqrt{(x_i - x_c)^2 + (y_i - y_c)^2} \tag{5}$$

$$\theta_i = arctan((y_i - y_c)/(x_i - x_c)) \tag{6}$$

The second constraint included in this colorogram model is color information, which is affected by parameters such as the number of bins and the method of similarity check. A true image with 24-bit RGB has got 16 million colors in it with red, green and blue channels accounting for 256 variations each varying from 0 to 255. Thus, 256 * 256 * 256 equalling 16,777,216 colors (~16 million) is available in a true image. The human eye cannot distinguish all these 16 million colors. This inability of the human eye is an advantage that has been exploited in reducing the number of bins. It is practically impossible to take all 16 million colors as bins and check for similarity. So the reduction of bins is required for effective computation.

$$c_{index} = (R/32 + 8\,(G/32) + 64\,(B/32)) \tag{7}$$

where C_{index} is the index value of the histogram bin that gives the color of the pixel.

Illustration of grouping of colors: An example is illustrated for the motivation to reduce the number of bins where a white pixel and a black pixel is considered. As the white pixel's red, green and blue components are all 255, its value of the colorogram index is 511. In contrary, a black pixel's red, green and blue components are all 0 yielding colorogram index value as 0. As a whole, the value of the index varies from 0 to 511 yielding 512 bins. This implies that in red, green and blue components; say 32 combinations of colors are grouped into one color. 32 * 32 * 32 colors are grouped as one color and assigned a bin in the colorogram. Thus, the 16 million colors have been reduced to 512 colors equaling 512 bins.

In HSV color space, the color model is reduced to 162 histogram bins, with hue constituting 18 levels, saturation and value yielding 3 levels each. For efficiency, the RGB color space is converted to HSV color space using the equations in (x). The bins in HSV color space is filled with respect to spatial constraint as in RGB space.

$$H = cos^{-1}\left\{\frac{\frac{1}{2}[(R-G)+(R-B)]}{\sqrt{(R-G)^2+(R-B)(G-B)}}\right\},$$

$$S = 1 - \frac{3}{R+G+B}[min(R,\ G,\ B)],$$

$$V = \frac{1}{3}(R+G+B) \tag{8}$$

Once the histograms are created, the histograms are compared to find their distance between each other for similarity. For that purpose, two histogram matching formulas are used in the form of Euclidean and Intersection based approaches. The intersection of histograms H_1 and H_2 is given by:

$$d(H_1, H_2) = \frac{\sum_A \sum_B \sum_C min(H_1(a,b,c), H_2(a,b,c))}{min(|H_1|, |H_2|)} \tag{9}$$

where $|H_1|$ and $|H_2|$ give the magnitude of each histogram, which is equal to the number of samples. The sum is normalized by the histogram with fewest samples.

The Euclidean distance between the color histograms H_1 and H_2 can be computed as:

$$d^2(h,g) = \sum_A \sum_B \sum_C min(h(a,b,c) - g(a,b,c))^2 \tag{10}$$

In this distance formula, there is only comparison between the identical bins in the respective histograms. Two different bins may represent perceptually similar colors but are not compared crosswise. All bins contribute equally to the distance.

3.2 Identity Tracker

The identity tracker identifies the blobs and tracks it across all frames. This module helps in solving the issues like correspondence of blobs between frames. Tracking is done through the colorogram model that is created on the fly in real time as

Algorithm 2: Labeling of Blobs

Input : Colorogram for a blob
Output : Tracked blobs with its trajectory

1. Correlate the colorograms of blob to all models.
2. Find maximum value of the correlation results.
3. If max. correlation value is less than a threshold
   Add a new model to the available list
   Assign colorogram of the blob to the new model
   Else
   Assign blob to model with max correlation value.
4. Compute centroid of blob for trajectory.
5. Output Tracked blobs

Fig. 2 Algorithm for labeling of blobs

explained in Fig. 2. The colorogram model of the blob establishes the correspondence of the blobs in consecutive frames. The centroid of the blob is computed and is stored across all frames for offline processing providing object's trajectory. This colorogram model even address the issue of tracking even under occlusion provided at least half of the object is visible. When the person enters the scene after some time he is correctly identified. So this colorogram approach helps in establishing motion correspondence as well as tracking under occlusion.

3.3 Human Blob Labeling

The objective of this paper is to label and track each individual consistently throughout the video sequence. The first problem to be solved when tracking people is detecting them. For detection of humans, an adaptive background subtraction method is used. Although the idea behind background subtraction is simple, the problem is very challenging. Changes in illumination conditions, movements of tree branches waving in the wind, camera noise, shadows or reflections can affect the detection performance. After this process, the system expects each person to be segmented into an isolated blob. Thus, the system assumes that in an indoor environment each detected blob will correspond to a person. As soon as a person enters the scene, a model for that person is generated and stored. A model consists of colorogram information of how the colors are distributed in the image. In the next frame, another model is built for each of the foreground blobs that are detected. Then the similarity measure between colorograms of stored models and blobs in the current frame are calculated using distance measures in (9) and (10). The similarity check is done using distance measures and the most similar blobs are matched to the model if their similarity measure is above an experimentally chosen threshold. However, if a blob in the current frame is at a significant distance from all the stored models, a new model is initialized. The fixing of threshold in any operation in images is still a mystery to the researchers. A local threshold and a global threshold are initialized at the start of labeling and these are updated dynamically with respect to the scene. Once the blobs are labeled, all the models are updated using an adaptive learning parameter 'α' as mentioned in (11).

$$h_I(c_i, t) = \alpha h_I(c_i, t-1) + (1-\alpha) h_I^{new}(c_i, t) \tag{11}$$

Once the labeling and updating of the models are done the next frame is processed and the process continues until all blobs are labeled in the video sequence.

4 Experiments and Analysis

4.1 Experimental Assumptions and Settings

The experiments were done on a system running Windows 7 OS powered with Intel Core i3 processor with 4 GB RAM. The identity tracker module was programmed with C using Intel's open source image library OpenCV. The test sequences consist of four persons with different clothing at an indoor environment. The four test subjects took their turns in entering the scene, doing arbitrary work and exit the scene. The test subjects were informed that random entry/exit of the scene and re-entry is mandatory and the frequency and ordering of their entry/exit/re-entry procedures is subjective. The identity tracker module is started with initialization of parameters like threshold, learning parameter (α) and the system labeled the human blobs. The colorogram model correctly identified the people as different persons entering the scene. Various correlation distance measures had been tested with these sequences and the results are attached with this section and the effects of the colorogram bin sizes are also discussed. More the number of bins, the sparser the array becomes. This is synonymous to more range in diversity of color, more complex computation involved. There exists a trade-off between the computational complexity and the accuracy of the color dispersion into bins. From the experiments conducted it is observed that the persons are correctly identified and the false positives are low. The appearance model of how each person's clothing appearance is closer to each other. The numbers in the cells indicate the level of closeness in appearance with respect to clothing. Higher the number in the cell, higher is the closeness. The values of the cells are obtained by correlating the colorograms of the blob with that of all available models. It is observed that Subject 1 stood out in appearance due to his strange orange color clothing that is not present with any other subjects. This is evident with the correlation results obtained.

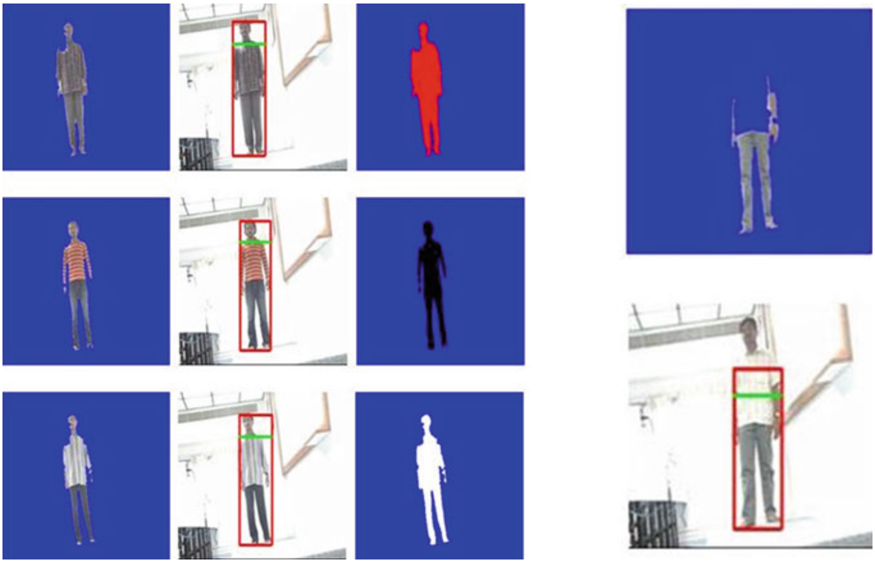


Fig. 3 Results of identity tracker for test subjects (*first row to third Row*) (*Column* 1) blobs extracted using adaptive background subtraction method (*Column* 2) subjects detected with bounding box and head detector (*Column* 3) Persons identified (*Labeled blobs*) with a color model (*Column* 4) person identified wrong due to error in background subtraction

The first column in Fig. 3 gives the color blob detected from the motion detector using an adaptive background subtraction. The bounding box enclosing the blob in red and the head tracker with a green horizontal line is shown in the second column. The third column shows the output of the identity tracker where each person is labeled different from each other with different colors used from the lookup (Red, Black, White, Green, Yellow, Cyan, Magenta). This coloring is synonymous with different labels and it is mainly used for visual appeal and ease of understanding. It is also observed that the maximum similarity is obtained when histogram is self-correlated with itself. When the first person enters the scene he will be labeled in red color and whenever he re-appears in the scene, he will be colored the same red color signifying that he was the same person present in the model and now he is re-entering the scene. And the next new person is identified and colored as black and so on.

4.2 Issues Encountered

Several issues were encountered during this process of human labeling. For instance, if there is a part of the body that has been segmented into a different blob after the background subtraction, and this blob has not been removed during the thresholding process, the system tends to fail by detecting that blob as a new person. It is also assumed that when a person enters the scene he/she will be isolated. If two or more people enter together and are segmented into the same blob by the background subtraction algorithm, the system will treat them as though they were a single person until they split. At this point the system will begin to keep track of each individual separately, assigning the label of the group to one of the components of the group and new labels for the rest. There would be the possibility for the system to go back in time and segment the people while they were together, but this was not implemented. There are very few scenarios where the algorithm fails miserably in identifying the right person. Figure 3 explains a scenario that is an effect of accuracy of the blob extracted. This shows the importance of the pre-processing stage. No efficient colorogram can be constructed if the person is wearing same colored clothing as that of background. If a test subject is wearing a top that is close in appearance with that of the background, then his torso will not be extracted due to the fact that his upper half getting subtracted in background subtraction stage. Therefore, only his head and leg can be extracted, not as single blob, but two separate blobs. It is assumed to consider blobs for computation if and only if they are above an empirically chosen size threshold value to avoid noise appearing as small-sized blobs. As this head blob for this person is small enough to be ignored; only the color information in the leg blob is taken for colorogram construction and the results can be seen in Fig. 3. As there is a huge amount of error in construction of colorogram and labeling, geometric localization of head with respect to torso and legs is devised. This localization procedure fits the head probable's with those of leg and torso probable's. This system cannot be applicable in places where people follow the principle of uniform clothing like schools, colleges, hospitals, etc.

The Precision-Recall curves, as in Fig. 4, for (1) Colorogram—Intersection and Euclidean (HSV & RGB), (2) Histogram and (3) Spatiogram methods are shown. From the curves, it is clearly evident that the HSV Intersection based Colorogram model yields the best results when compared with other approaches. Only 5 % false positives are present with these four test subjects. To compare different approaches, dice coefficient [10] is used. Given the ground truth for a blob Ω_1 and tracked region Ω_2, the Dice Coefficient is given by,

$$D(\Omega_1, \Omega_2) = \frac{2\,Area(\Omega_1 \cap \Omega_2)}{Area\,(\Omega_1) + Area(\Omega_2)} \tag{12}$$

Dice coefficient D varies from 0 to 1. This coefficient is computed for all the three approaches and a graph showing the results are given in Fig. 5. It is inferred from the graph that Colorograms perform on par with Spatiograms but clearly outperforms Histograms. This is because of the fact both spatiograms and colorograms includes or captures geometric information along with color information from the scene.

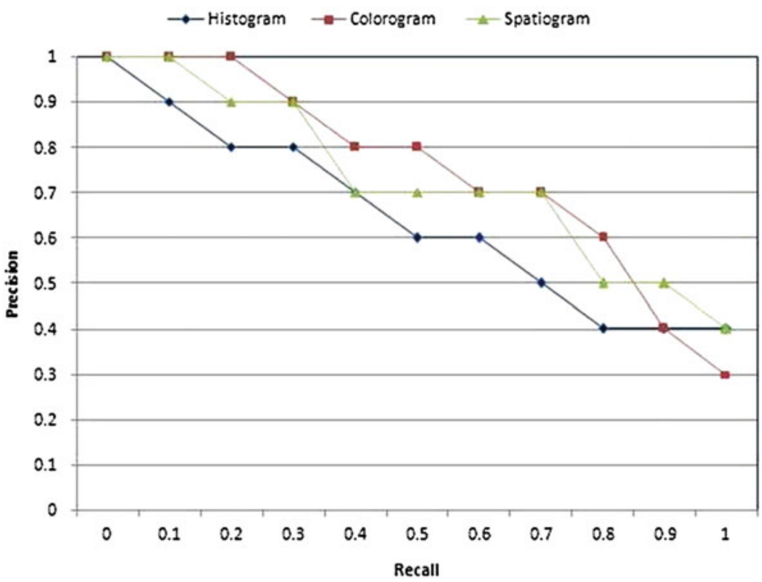
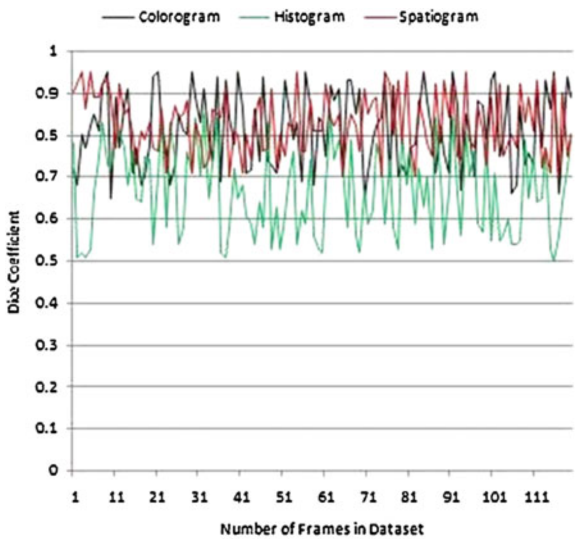


Fig. 4 Precision recall curves for histogram, spatiogram and colorogram

Fig. 5 Dice coefficients graphs for histogram, spatiogram and colorogram



5 Conclusion and Future Work

Various issues were encountered during this module and some of the issues were addressed fully and some partially. At first, thresholding is considered to be one of the main problems in this module while correlation of colorograms is done.

An empirical way of fixing the thresholds is tried and failed. Then an experimentally chosen threshold is fixed to solve this problem. Secondly, the issues during entry/exit of persons are taken care of by assuming that the blobs need to be processed only when it is away from the image boundary fully. In spite of all these issues, this module works satisfactorily when people enter the scene with different clothing. When feature like shape of the contour is taken along with the color information, the system will improve in efficiency and it discriminates the people wearing same dress. Spatiograms [11,12] and Histograms were compared and reasonably good results were obtained. And, intersection based histogram matching proved its superiority over Euclidean distance measure. Adding Gait information to this colorogram model will identify the person along with his clothing information effectively.

References

1. Elgammal A, Harwood D, Davis L (2000) Non-parametric model for background subtraction. In: Proceedings of the 6th European conference on computer vision (ECCV 2000). Dublin, Ireland, Springer, pp 751–767
2. Lipton AJ, Fujiyoshi H, Patil RS (1998) Moving target classification and tracking from real-time video. In: Proceedings of the DARPA image understanding workshop (IUW'98), November 1998. Monterey, USA, pp 129–136
3. Desa SM, Salih QA (2004) Image subtraction in real time moving object extraction. In: Proceedings of international conference on computer graphics, imaging and visualization, January 2004
4. Capellades MB, DeMenthon D, Doermann D (2004) An appearance based approach for consistent labeling of humans and objects in video. In: Proceedings of the pattern analysis and applications, November 2004. pp 1433–7541
5. Chleq N, Br′emond F, Thonnat M (1998) Image understanding for prevention of vandalism in metro stations. In: Proceedings of advanced video-based surveillance systems, November 1998, vol 488. USA, pp 106–116
6. Haritaoglu I, Harwood D, Davis LS (2000) W4: real-time surveillance of people and their actions. IEEE Trans Pattern Anal Mach Intell 22(8):809–830
7. Alamdar F, Keyvanpour M (2011) A new color feature extraction method based on quad histogram. In: Proceedings of the 3rd international conference on environmental science and information application technology, January 2011, vol 10. pp 777–783
8. Yin X, Duntsch I, Gediga G (2011) Quad tree representation and compression of spatial data. Lect Notes Comput Sci 6499:207–239
9. Birchfield ST, Rangarajan S (2005) Spatiograms versus histograms for region-based tracking. In: Proceedings of the IEEE conference on computer vision pattern recognition, vol 2. pp 1158–1163
10. Boesen K et al. (2004) Quantitative comparisons of four brain extraction algorithms. Neuro Image 22(3):1255–1261
11. Br′emond F, Thonnat M (1997) Tracking multiple non-rigid objects in a cluttered scene. In: Proceedings of the 10th Scandinavian conference on image analysis (SCIA'97), June 1997, vol. 2. Lappeenranta, Finland, pp 643–650
12. Martinez-Uso A, Salgues G, Velastin SA (2011) Evaluation of unsupervised segmentation algorithms for silhouette extraction in human action video sequences. Lect Notes Comput Sci 7066:13–22

ECC Based Biometric Encryption of Compressed Image for Security over Network Channels

B. Prasanalakshmi and A. Kannammal

Abstract This paper provides a way for secure transmission of data over networks. The security is initialized by means of using biometric data as private key, imposing the biometric data compression using 5- level Discrete wavelet transform is a way to minimize the key length, which were of large length as the standard requirements for implementing via RSA based hardware and software. As additional security measure Elliptic curve cryptography is used to encrypt the biometric data that is to be stored in and accessed from the database over networks.

Keywords 5-DWT · ECC · Biometrics · Encryption · Cryptography

1 Introduction

In the era of computers and networks, much of the communications between people is facilitated by networks. During this communication both the parties are really worried about their data confidentiality, data authentication non repudiation and privacy etc. That too when the data is concerned with that of government agencies, then confidentiality level is high. In order to bring security two well-known kinds of cryptographic protocols namely public key and symmetric key protocol.

B. Prasanalakshmi (✉)
Department of CSE, Thirumalai Engineering College, Kanchipuram, India
e-mail: bplakshmi@ieee.org

A. Kannammal
Coimbatore Institute of Technology, Coimbatore, India
e-mail: kannaphd@yahoo.co.in

Mohan S. and S. S. Kumar (eds.), *Proceedings of the Fourth International Conference on Signal and Image Processing 2012 (ICSIP 2012)*, Lecture Notes in Electrical Engineering 222,
DOI: 10.1007/978-81-322-1000-9_33,

In symmetric key protocol such as DES, IDEA and AES, a common key is used by both sender and receiver for encryption and decryption. This system provides high speed but have the drawback that a common key must be established for each pair of participants. In public key protocol there are two keys, public key and private key by which message can be encrypted and decrypted. One is kept private by owner and used for decryption. The other key is published to be used for encryption. Some of the most useful example of the public key cryptography is RSA, ElGamal and DSA [1]. Although these systems are slower but they provide arbitrary high level security. Due to comparative slowness of the public key cryptography algorithms, dedicated hardware support is desirable. Most of the network and standards that uses public key cryptography for encryption and digital signature uses RSA. The key length for secure RSA has increased over recent years, and this is putting a heavier load on application of RSA. Thus it creates extra computation cost and processing overhead. Thus in this paper we have used Elliptic Curve Cryptography (ECC). The principal attraction of ECC compared to RSA is that it offers higher security per bit with smaller key size. It provides higher security per bit. Since elliptic curve cryptography has smaller key size, hence it also reduced the computation power, memory and bandwidth. This paper is organized as follows. In Sect. 2, we provide the review of the elliptic curve cryptography, why we use elliptic curve cryptography instead of RSA or other cryptography system, the implementation method of ECC and its mathematical operation and method for finding all points on the elliptic curve on which we have to encrypt the message. We describe the Elliptic Curve Diffie-Hellman Algorithm (ECDH) in this section for generating key. Section 3 discusses the previous works. In Sect. 4 we describe how can we encrypt and decrypt the message by the help of palm vein as a private key. In Sect. 5 we describe the merits of the suggested approach. We conclude the paper in Sect. 6.

2 Elliptic Curve Cryptosystem

In 1985, Neil Koblitz and Victor Miller independently proposed the use of elliptic curve cryptography. Since 1985, there have been a lot of studies concerning elliptic curve cryptography. The use of ECC is very inviting for various reasons [1, 2]. The first and probably most important reason is that ECC offers better security with a shorter key length than any other public-key cryptography. For example, the level of security achieved with ECC using a 160-bit key is equivalent to conventional public key cryptography (e.g. RSA) using a 1024-bit key [1]. There is huge importance of shorter key lengths especially in applications having limited memory resources because shorter key length requires less memory for key storage purpose. Elliptic curve cryptosystems also require less hardware resources than conventional public-key cryptography. Now at the security level ECC is more secure than RSA. RSA can be cracked successfully, uses 512 bits and for ECC the number of bits is 97, respectively. It has been analysed that the computation power required for cracking ECC is approximately twice the power required for cracking RSA. ECC provides higher level of security due to its complex mathematical operation. Mathematics used for ECC is considerably more difficult and deeper than mathematics used for conventional cryptography. In fact this is the main reason, why elliptic curves are so good for cryptographic purposes, but it also means that in order to implement ECC more understanding of mathematics is required. For any operation on elliptic curve, first of all we have to find the all point of that curve. Thus for finding the point on the curve firstly we have to choose any elliptic curve. Suppose $y_2 \bmod p = (x^3 + ax + b) \bmod p$ is an elliptic curve, Where $4a^3 + 27b^2 \neq 0$. Then points on this curve are the set E_p (a, b) consisting of all pairs of integers (x, y), which satisfy the above equation together with the point Zero.

Method for finding the points on the curve is as follows.

Step 1. Determine the L.H.S of elliptic curve For all (x,y) € Z_p
Step 2. Determine the R.H.S of elliptic curve For all x,y € Z_p
Step 3. Choose the Pair of corresponding value of x and y as a pair for all x, y € Zp for which L.H.S. = R.H.S
Step 4. All pairs of such (x, y) are the point on the curve

Example

Elliptic Curve Parameters: P = 751, a = −1, b = 188, Generator Point G = (297,569), nA = 13, nB = 12. Followings are the encrypted coordinates of the ciphered session key : [(45, 97), (64, 738), (333, 435), (333, 435), (324, 7), (45, 97), (84, 613), (529, 254), (653, 422), (414, 88), (45, 97), (492, 167), (627, 59), (472, 336), (627, 59), (297, 569), (160, 140), (84, 613), (472, 336), (492, 167), (627, 59), (492, 167), (732, 180), (324, 7), (653, 422), (45, 97), (333, 435), (653, 422), (297, 569), (297, 569), (297, 569), (529, 254), (160, 140), (160, 140), (653, 422), (653, 422), (627, 59)] point on elliptic.

2.1 Elliptic Curve Diffie-Hellman Algorithm

Elliptic curve Diffie-Hellman algorithm is the Diffie-Hellman algorithm for the elliptic curve [1]. The original Diffie-Hellman algorithm is based on the multiplicative group modulo *p*, while the elliptic curve Diffie-Hellman (ECDH) protocol is based on the additive elliptic curve group. We assume that the underlying field *GF* (*p*) is selected and the curve *E* with parameters *a*, *b*, and the Base point *P* is set up. The order of the base point *P* is equal to *n*. The standards often suggest that we select an elliptic curve with prime order, and therefore, any element of the group would be selected and their order will be the prime number *n*. At the end of the protocol the communicating parties end up with the same value *K*.

Which is a point on the curve. A part of this value can be used as a secret key to a secret-key encryption algorithm. Suppose there are two users Alice and Bob. According to the Diffie-Hellman the key generation and exchange is as follows.

Step 1. Alice chooses her elliptic curve and two points
Step 2. Alice also chooses his secret value d_A which is a random number
Step 3. Now suppose Alice curve parameter is (a, b), prime number is P
Step 4. Now Alice compute the $P_a = d_A{*}G$
Step 5. Now Pa is the public key of the Alice
Step 6. Bob gets the Alice's public key and computes his own d_A which is also a random number
Step 7. Now Bob multiply d_b with G and compute $P_b = d_b{*}\ G$
Step 8. This P_b is the bob public key
Step 9. Now Bob then multiply $d_b{*}P_a$ and gets the secret key, i.e. $K = d_b{*}P_a$
Step 10. Similarly Alice multiply his own private key with Bob's public value and gets the secret key. i.e. $K = d_a{*}P_b$
Step 11. Thus secret key is $K = d_a{*}P_b = d_b{*}P_a = d_a{*}d_b{*}G$;
Step 12. By exchanging the key through this method both Bob and Alice can communicate safely. Bob can use the secret value he computed to build an encrypting key

When Alice gets the message from Bob, she use the secret value she computed to build the decrypting key. It's the same secret value, so they use the same key. Thus what Bob encrypts Alice can decrypt.

2.2 Palm Vein

If an individual has registered his profile in his childhood, it'll still be recognized as he grows, as an individual's patterns of veins are established *in utero* (before birth) and No two people in the world share a palm vein pattern—even those of identical twins differs it is specified [3] as much reliable. The discussion also proceeds from the pre-processing steps of palm vein feature extraction till the generation of key. The features of Palm vein extracted are basically Minutiae points like bifurcations and edges. These minutiae points are extracted using the methods discussed. Then an irrevocable key is generated from the cancellable palm vein template using the proposed method [4].

A typical vein pattern biometric system consists of four individual processing steps: Image Acquisition, Image Enhancement, Vein Pattern Segmentation and Feature Extraction as shown in Fig. 1. During image Acquisition the images are usually captured using infra-red imaging technologies. One of the methods uses a Far—Infrared (FIR) camera to acquire the thermal vein pattern images of the palm. After obtaining the images, the system segments the vein pattern from the background and extracts the feature to obtain the shape, representation of the pattern. Finally the system recognises the vein pattern using various pattern recognition methods. Analysing the database of infrared vein patterns shows

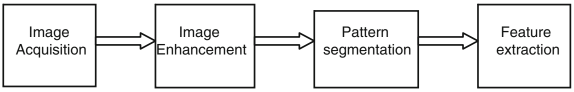


Fig. 1 Steps in vein feature recognition

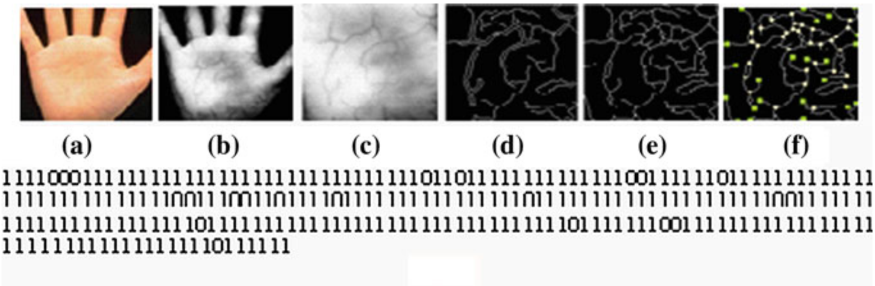


Fig. 2 Steps in key generation **a** raw image, **b** smoothened image, **c** region of interest, **d** segmented image, **e** thinned image, **f** extracted minutiae points, **g** generated key

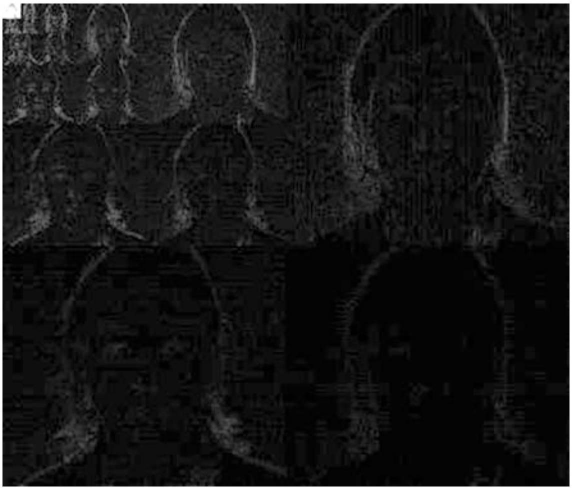
a trend that for each palm vein pattern image there are on average 13 minutiae points in each vein pattern image, including 7 bifurcation and 6 Ending points. Figure 2 shows the key generated from the palm vein [4].

3 Previous Work

The main problem of asymmetric cryptography is the management of private key. No one should be able to access someone else's private key. They need to store in such a place which is protected from unauthorized accessing. This is vulnerable to attack of hackers. This creates big problem in asymmetric cryptography. Thus it can be solved by the use of biometric template. Private Key can be generated directly by the biometric template. Since private key can be generated dynamically from one's biometric template, so there is no need to store private key anymore and network becomes more secure and safe. But there are very little work has been done in the field of elliptic curve cryptosystem with the help of biometric. One of the suggested approaches is given [2].

However these biometric has lots of issue regarding training, capturing image, obscurity by eyelash, eyelids etc. from the camera, lack of existing data deters ability, cost etc. For some individuals, the iris image capturing is very difficult. Iris recognition system requires lots of memory to be stored. It is easily absured by eyelash, eyelids, lens and reflectness from the cornea. People are not much familiar with iris recognition system yet, so there are lots of myths and fears related to scanning the eye with light source. Iris recognition system works on the basis of acquisition of iris image, but acquisition of an iris image needs more training and

Fig. 3 5- Level DWT



attractiveness than most biometrics. It also cannot verify by human. The most problem with iris recognition system is its expensiveness. When we compare all biometric then we show that palm vein is the most adequate methodology for authentication. Hence in proposed method we have used palm vein as a private key instead of other biometric.

4 Proposed Work

In this paper we use palm vein and cryptographic hash function to generate private key.

4.1 Encoded Image Compression Using 5-DWT

Image compression is carried out using DCT and then performing 5 level DWT and the secret image encoding is done using Anti- Arnold transformation as used in the previous works [5]. 5- level DWT of the encoded face image is shown as in Fig. 3.

4.2 Method for Generating Public Key and Private Key

To generate private key, we take the palm vein of the user and generate its hash value by the help of MD5 cryptographic hash function [6]. This resultant hash value is the private key of the user. Suppose this value is d_a for use A and d_b for user B. Now to generate public key in elliptic curve cryptosystem with the help of this private key the following procedure is followed :-

Step 1. Both user choose the same large prime 'p' and the elliptic curve parameter 'a' and 'b' such that y_2 mod p = (x_3 + ax + b)mod p, Where $4a^3 + 27b^2 \neq 0$

Step 2. Now choose any one point G(x,y) from this elliptic curve. This point is called the base point of the curve

Step 3. Compute $P_a = d_a$ *G(x,y) This P_a is called the public key of user A

To generate public key of user B same operation can be performed by the help with private key of user B

4.3 Message Encryption

Suppose user 'A' wants to send a message to user 'B'. Then first task in this system to encode the 5 level compressed image to be sent as a point P_m (x, y). Encoding of the compressed image is done using Anti-Arnold transformation. It is the point P_m that will be encrypted as a cipher image and subsequently decrypted. After mapping of points with the feature points on elliptic curve, they can encrypt the message by following steps.

Step 1. Suppose User A encodes the message m as P_m = (x,y)

Step 2. User A take his private key from his palm vein suppose it is k

Step 3. User A compute the k*G

Step 4. User A compute the P_m + k*P_b here P_b is the public key of user B and Pm is the face image

Step 5. User A take the Cm = (k*G, Pm + k*P_b) as a cipher image

4.4 Message Decryption

For Message decryption we have to do following procedures.

Step 1. User B takes the first point of the encrypted face data

Step 2. User B now compute d_b*k*G

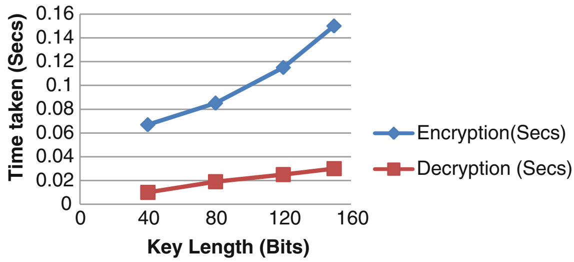
Step 3. User B then subtracts it with his second point

Step 4. Thus user B compute P_m + k*Pb-db*k*G = P_m

Step 5. The image P_m is the required image data of User B which is sent by User A

Step 6. User A can send this cipher image to User B

Fig. 4 Key length versus time taken -ECC



5 Merits of Proposed Approach

Traditional methods for implementing public key infrastructure and encryption and decryption techniques faces lots of problem such as key management, key storing, key privacy etc. Our proposed approach can handle such problems. Here we are using palm vein as a private key so that there is no need to store any private key and also palm vein has lots of merits over other biometrics like it is most user friendly and cheaper. Palm vein recognition also has some outstanding features like universality, permanence, uniqueness and accuracy. As we are using elliptic curve cryptosystem, so we can achieve high level security with very shorter key size. Thus it also solves the key size problem. As we know that ECC requires very complex mathematical operation (because of elliptic curve Diffie-Hellman problem, which is harder than discrete logarithmic problem) therefore security strength per bit is also very high. Also the usage of the encoded and watermarked biometric image gives much security so that image may not be tampered or not be used even on theft.

6 Results and Discussion

Time taken for encoded watermarked image Encryption and Decryption is provided in the Fig. 4. This shows improved efficiency among the existing systems.

7 Conclusion

In this paper, communication between more than one network becomes very secure with the help of elliptic curve and palm vein. The main advantage is that it requires very less key size and gives high level of security with palm vein recognition system and there is no need to store any private key anywhere. Thus it is more efficient and highly secure using highly accurate biometric.

References

1. Mel HX, Baker D (2011) Cryptography decrypted. Addision- Wesley
2. Mohammadi S, Abedi S (2008) ECC based biometric signature : a new approach in electronic banking security. In: international symposium on electronic commerce and security (ISECS'07), pp 763–766, doi:10.1109/ISECS.2008.98
3. Vein pattern recognition: www.fujitsu.com
4. Prasanalakshmi B, Kannammal A (2009) A secure cryptosystem from palm vein biometrics. In: Proceedings of the 2nd international conference on interaction sciences: information technology, culture and human (ICIS '09). ACM, New York, pp 1401–1405. doi:10.1145/1655925.1656183
5. Prasanalakshmi B, Kannammal A (2011) Frequency domain combination for preserving data in space specified token with high security. In: Communications in computer and information science, Springer, Berlin, pp 319–330, vol 251. doi:10.1007/978-3-642-25327-0_28
6. Anoop MS (2008) Elliptic curve cryptography, an implementation tutorial, Tata Elexsi Ltd, Thiruvananthapuram, India
7. Waang L, Leedham CG (2005) A thermal vein pattern verification systems, pattern recognition and image analysis. Lecture notes in computer science, vol 3687, Springer, Berlin, pp 58–65

Video Shot Detection Using Cumulative Colour Histogram

B. H. Shekar, K. Raghurama Holla and M. Sharmila Kumari

Abstract The shot boundary detection is the fundamental step in video indexing and retrieval. In this paper a new method for video shot boundary detection based on *slope* and *y-intercept* parameters of the straight line fitted to the cumulative plot of color histogram is proposed. These feature vectors are extracted from every video frames and the frame dissimilarity values are compared against a threshold to identify the *cuts* and *fades* present in the video sequence. Experiments have been conducted on TRECVID video database to evaluate the effectiveness of the proposed model. A comparative analysis with other models is also provided to reveal the superiority of the proposed model for shot detection.

Keywords Cumulative colour histogram · Cut detection · Fade detection · Video segmentation · Shot boundary detection

B. H. Shekar (✉) · K. Raghurama Holla
Department of Computer Science, Mangalore University,
Konaje, Mangalore, Karnataka, India
e-mail: bhshekar@gmail.com

K. Raghurama Holla
e-mail: raghu247@gmail.com

M. Sharmila Kumari
Department of Computer Science and Engineering, P. A. College of Engineering,
Mangalore, Karnataka, India
e-mail: sharmilabp@gmail.com

Mohan S. and S. S. Kumar (eds.), *Proceedings of the Fourth International Conference on Signal and Image Processing 2012 (ICSIP 2012)*, Lecture Notes in Electrical Engineering 222,
DOI: 10.1007/978-81-322-1000-9_34,

1 Introduction

As a consequence of rapid development in video technology and consumer electronics, the digital video has become an important part of many applications such as distance learning, advertising, electronic publishing, broadcasting, security, video-on-demand and so on, where it becomes more and more necessary to support users with powerful and easy-to-use tools for searching, browsing and retrieving media information. Since video databases are large in size, they need to be efficiently organized for quick access and retrieval. Shot detection is the fundamental step in video database indexing and retrieval. The video database is segmented into basic components called shots, which can be defined as an unbroken sequence of frames captured by one camera in a single continuous action in time and space. In shot boundary detection, the aim is to identify the boundaries by computing and comparing similarity or difference between adjacent frames. So, the video shot detection provides a basis for video segmentation and abstraction methods [7].

In general, shot boundaries can be broadly classified into two categories [3, 7]: abrupt shot transitions or *cuts*, which occur in a single frame where a frame from one shot is followed by a frame from a different shot, and gradual shot transitions such as *fades*, *wipes*, and *dissolves*, which are spread over multiple frames. A *fade-in* is a gradual increase in intensity starting from a solid color. A *fade-out* is a slow decrease in brightness resulting in a black frame. A *dissolve* is a gradual transition from one scene to another, in which the two shots overlap for a period of time. Gradual transitions are more difficult to detect than abrupt transitions.

The remaining part of the paper is organized as follows. Section 2 presents the review of existing models for shot detection. The proposed model is discussed in Sect. 3. Experimental results and comparison with other models are presented in Sect. 4 and conclusion is provided in Sect. 5.

2 Review of Existing Work

There are several models available in the literature for video shot boundary detection. To identify shot boundary, the low level visual features are extracted and the inter-frame difference is calculated using these extracted features. Each time the frame difference measure exceeds the threshold, the presence of shot boundary is declared. Some of the popular existing models for cut detection and fade detection are reported in the following section.

2.1 Cut Detection: A Review

The simplest way of detecting the hard cut is based on pair-wise pixel comparison [3], where the differences of corresponding pixels in two successive frames are computed to find the total number of pixels that are changed between two consecutive frames. The absolute sum of pixel differences is calculated and then compared against a threshold to determine the presence of cut. However, this method is sensitive to object and camera motions. So an improved approach is proposed based on the use of a predefined threshold to determine the percentage of pixels that are changed [22]. This percentage is compared with a second threshold to determine the shot boundary.

Since histograms are invariant to rotation, histogram based approaches are used in [22]. Whenever the histogram difference between two consecutive frames exceeds some threshold, a shot boundary is detected. When the color images are considered, some appropriate weights are assigned to the histogram of each color component depending on the importance of color space, so weighted histogram based comparison was suggested [9].

To improve the accuracy of shot identification process, low level visual features such as edges and their properties are used in some of the techniques. In [21] proposed a shot detection method by analysing *edge change ratio* (ECR) between consecutive frames. The percentage of edge pixels that enter and exit between two consecutive frames are calculated. The cuts and gradual transitions can be detected by comparing the ratio of entering and exiting edge pixels. The limitation of the edge based methods is that they fail to produce good results when the video sequence contains high speed object motions. To overcome this shortcoming, an algorithm for motion compensation was developed [2, 22]. The block matching procedure is used and motion estimation is performed. However, motion based approach is computationally expensive.

There are some models which use statistical features for shot detection [2, 9]. The image is segmented into a fixed number of blocks and the statistical measures like mean and standard deviation of pixels in the blocks are computed. These features are extracted from successive frames and compared against a threshold to detect a shot cut. As this type of methods may lead to false hits, more robust techniques using likelihood ratio (LHR) was proposed [9]. Even though likelihood ratio based methods produce better results, extracting statistical features is computationally expensive.

In [10] proposed a text segmentation based approach for shot detection. The combination of various features such as, color moments and edge direction histogram are extracted from every frame and distance metrics are employed to identify shots in a video scene. Gabor filter based approach was used for cut detection by convolving each video frame with a bank of Gabor filters corresponding to different orientations [17].

In [13], dominant color features in the HSV color space are identified and a histogram is constructed. Block wise histogram differences are calculated and the shot boundaries are detected. In [6], a shot segmentation method based on the concept of visual rhythm is proposed. The video sequence is viewed in three dimensions: two in the spatial coordinates and one in the temporal i.e. corresponding to frame sequence. The visual rhythm approach is used to represent the video in the form of 2D image by sampling the video. The topological and morphological tools are used to detect cuts. We can see that the combinations of various features such as pixels, histograms, motion features etc used to accurately detect shot transitions [12]. The frame difference values are computed separately for individual features to address different issues such as flash light effect, object or camera motion. For more review on shot boundary detection models see [15, 19].

2.2 Fade Detection: A Review

Once the abrupt transitions are identified, then it is very important to detect the gradual shot transitions such as *fade in* and *fade outs* which are present in the video for designing an effective video retrieval system. Usually, gradual transitions are hard to detect when compared to abrupt transitions.

In histogram based approaches, where the histogram differences of successive frames are computed and two thresholds [22] are used to detect the shot transitions: higher threshold T_h for detecting cuts and lower threshold T_l for detecting gradual transition. Initially, all the cuts are detected using T_h and then the threshold T_l is applied to the remaining frame difference values to determine the beginning of gradual transition. Whenever the start frame of the gradual transition is detected, the cumulative sum of consecutive frame difference is calculated until the sum exceeds T_h. Thus the end frame of the gradual transition is found. However, there is a possibility of false positives if the thresholds are not properly set.

In [1] suggested a fade detection model by analysing the first and second derivative of the luminance variance curves. This method has a limitation of being sensitive to motion and noise. Fernando et al. [4] used statistical features of luminance and chrominance components to detect fades. Zabih et al. [21] developed a model for shot detection based on analysing edge change ratio between adjacent frames. The percentage of edge pixels that enter and exit between consecutive frames are computed. The cuts and gradual transitions can be detected by comparing the ratio of entering and exiting edge pixels. Troung et al. [16] improves the fade detection using two step procedure: detecting the solid color frame in the first step and searching for all spikes in the second derivative curve. In another approach, fade in and fade out transitions were detected using histogram spans of video frame [5], which works based on the variations of the dynamic range during fade transitions. A fade detection method was proposed based on the principle that the horizontal span of histogram should increase for fade in and decrease for fade out. Each frame is divided into four regions and the histogram spans for these regions are computed to identify the beginning and end frame of fade transitions. In [6], Guimarães et al. proposed a shot detection algorithm based on Visual Rhythm Histogram (VRH). Fade regions are identified by detecting the inclined edges in VRH image. However, if the solid color frames are not solid black or white, then it results in two cross edges in the VRH image. In [18], localized edge blocks are used to identify gradual transition detection based on the concept of variance distribution of edge information in the frame sequences.

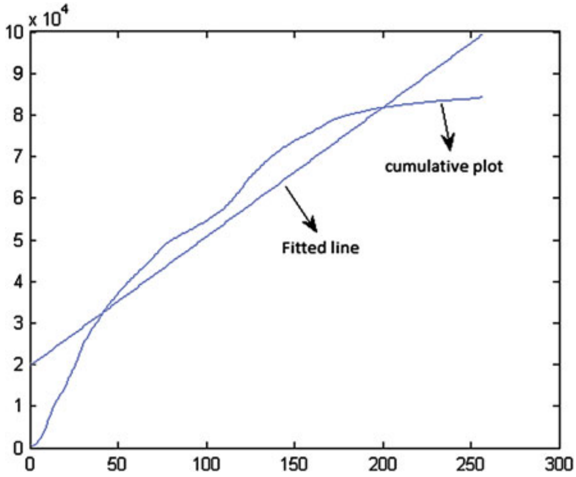


Fig. 1 Fitting a *straight line* to the cumulative plot

We have seen that several algorithms are available for shot boundary detection. The proposed technique is different from the existing models in terms of the reduced dimension of feature vector, hence suitable for real time video database processing applications. In the proposed work, we focussed on detecting both hard cuts and fade transitions present in the video. The details of the proposed model are given in Sect. 3.

3 Proposed Model

The proposed model is based on the *slope* and *y-intercept* parameters of the cumulative plot of the color histogram. Given a video containing n frames, the histograms are computed for all the three channels R, G and B separately. The cumulative histograms are obtained from each histograms. We fit a straight line to the cumulative histogram as shown in Fig. 1. Then the line parameters such as *slope* and *y-intercept* are calculated. So, it results in six element feature vector for each video frame. This feature vector is subsequently used to identify the shot transitions based on the frame differences. The frame difference between two consecutive frames is calculated as follows:

$$D(n) = d(F_n - F_{n-1}) \tag{1}$$

where

$$d(F_i - F_{i-1}) = \sum_{k=1}^{N} \|F_i(k) - F_{i-1}(k)\|^q \tag{2}$$

where N is the length of the feature vector, i.e. $N = 6$. When $q = 2$, $d(i,j)$ is the Euclidean distance between the frame i and frame j. The value of D indicates the change tendency of consecutive frames.

3.1 Cut Detection

The inter-frame dissimilarity values obtained in the previous stage are analysed to perform cut detection. We have used the *local adaptive thresholding* [20] since using a single global threshold is inadequate to detect all the cuts present in the video. In the first stage, global threshold T_g [11] is calculated using global mean μ_g and standard deviation σ_g obtained from the frame dissimilarity values D as follows:

$$T_g = \mu_g + \beta \cdot \sigma_g \tag{3}$$

where β is a constant which controls the tightness of T_g.

In the second stage, we consider only those frame dissimilarity values where, $D(V_n, V_{n-1}) \geq T_g$ for calculating the local threshold. Using a sliding window of size m, the mean and the standard deviation of left side frame difference(video frames: 1 ... $\frac{m}{2} - 1$) and right side frame difference(video frames: $\frac{m}{2} + 1$ ... m) from the middle frame difference$(m/2)$ are computed. The middle frame is declared to be the *cut* if the following conditions are completely satisfied:

(i) The middle frame difference value is the maximum within the window.
(ii) The middle frame difference value is greater than $max(\mu_L + \sigma_L T_d, \mu_R + \sigma_R T_d)$, where μ_L, σ_L and μ_R, σ_R represent mean and standard deviation values of the left side and right side of the middle frame difference value respectively. T_d is a weight parameter.

We have empirically fixed the values of β, m and T_d to be 1.5, 9 and 5 respectively.

3.2 Fade Detection

In order to detect fade transitions, we consider only the frame differences corresponding to non-cut frames, i.e. the fade-detection is followed by cut-detection process. Once the cuts are detected in a video segment, the associated frame difference values are not considered for further processing. The threshold selection for fade detection is done using local adaptive thresholding technique as suggested in [8]. A sliding window is placed preceding the current frame difference value. The mean (μ) and standard deviation (σ) within the window is calculated. The thresholds $\mu + \sigma k_1$ and $\mu + \sigma k_2$ are used to detect the beginning and end of fade transition respectively, where $k_1 = 2$ or 3 and $k_2 = 5$ or 6.

Fig. 2 Pair of consecutive frames showing cuts for the video segment *Lecture series*

4 Experimental Results

This section presents the experimental results to reveal the success of the proposed model. We have conducted experimentation on TRECVID video database as this is one of the standard database used by many researchers as a benchmark to verify the validity of their proposed shot detection algorithms. All experiments have been performed on a *Intel CORE* i5 processor, Windows operating system with 4 GB of RAM. We performed experiments with several videos from the database and obtained better results for all the videos. Some of the results are given in Sects. 4.1 and 4.2.

4.1 Cut Detection Results

We have considered the first 10,000 frames for the video segment *Lecture series*. Figure 2 shows the pair of consecutive frames corresponding to cuts for the first 5,000 frames. The cut detection result obtained from the proposed model is shown in Fig. 3a. The sharp peaks in Fig. 3a correspond to *cuts*. The proposed model accurately detected all the 27 cuts present in the video segment. The cut detection result for the first 2,500 frames of the video segment *Mechanical works* is shown in Fig. 3b. Figure 3c shows the result for the video segment *Central valley project*.

The performance of the proposed model is evaluated using precision and recall as evaluation metrics. The precision measure is defined as the ratio of number of correctly detected cuts to the sum of correctly detected and falsely detected cuts of a video data and recall is defined as the ratio of number of detected cuts to the sum of detected and undetected cuts. Also we compared our results by performing

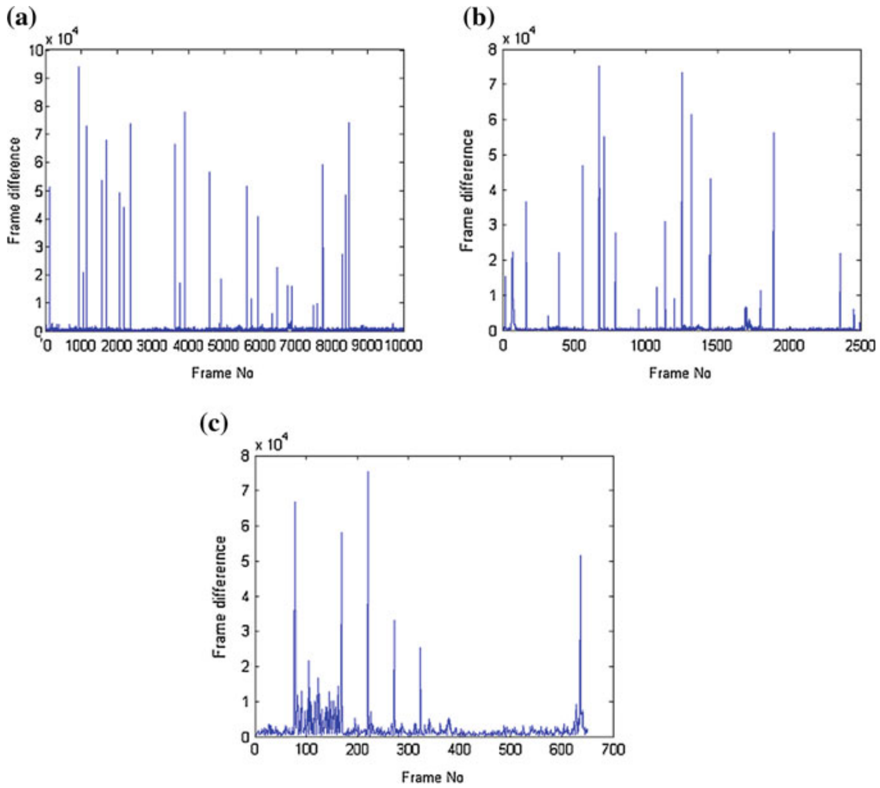


Fig. 3 Plot of *frame number* versus *frame dissimilarity measure* for the video segment: **a** *Lecture series*, **b** *Mechanical works*, **c** *Central valley project*

similar experimentation on the same video segments using other shot detection models based on *Pixel difference* [2], *Edge Change Ratio* (ECR) [21] and *chromaticity histogram* [14], and the results are reported in Table 1.

4.2 Fade Detection Results

The proposed model was tested with several videos containing fade transitions and produced satisfactory results. The fade transition regions for the *Tom and Jerry* video segment are shown in Fig. 4a. We considered the first 1,000 frames as they contained fade transitions. Figure 5 shows the frame numbers resulted from the experiment indicating the start and end of fade transitions. The fade detection result for the video segment *America's New Frontier* is shown in Fig. 4b.

The performance of the proposed model for fade detection is evaluated using precision and recall measures. These parameters were obtained for the proposed

Table 1 Precision and recall metrics of the proposed model for cut detection on TRECVID video segments

Video segment	Metrics	Proposed model	Pixel difference model [2]	ECR based model [9]	Chromaticity histogram [14]
Lecture series	Precision	1.00	0.96	0.78	1.00
	Recall	1.00	1.00	0.84	1.00
	F^1 value	**1.00**	**0.98**	**0.80**	**1.00**
Mechanical works	Precision	0.94	0.90	0.84	0.89
	Recall	0.89	1.00	0.88	0.89
	F^1 value	**0.91**	**0.95**	**0.86**	**0.89**
Central valley project	Precision	0.87	1.00	0.60	1.00
	Recall	1.00	0.50	0.43	0.75
	F^1 value	**0.93**	**0.67**	**0.50**	**0.86**

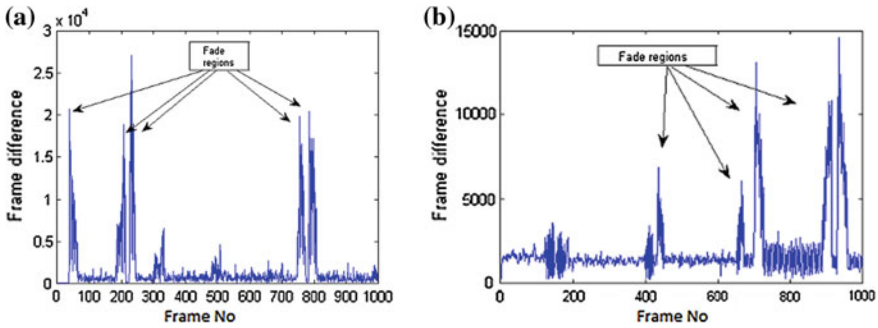


Fig. 4 Plot of *frame number* versus *frame dissimilarity measure* for the video segment: **a** *Tom and Jerry*, **b** *America's New Frontier*

Fig. 5 Pair of frames representing fade transition for the video segment *Tom and Jerry*

model on three different video segments. In order to compare our results, we have also conducted similar experimentation on the same video segments with another fade detection algorithm called *Twin comparison* method [22] and results of our model and *Twin comparison* based algorithm is reported in Table 2.

Table 2 Precision and recall metrics of the proposed model for fade detection on TRECVID video segments

Video segment	Metrics	Proposed model	Twin comparison based model
Tom and Jerry	Precision	1.000	1.000
	Recall	1.000	0.833
	F^1 measure	**1.000**	**0.909**
America's New Frontier	Precision	1.00	1.000
	Recall	1.000	0.909
	F^1 measure	**1.000**	**0.925**
Central Valley Project	Precision	1.000	1.000
	Recall	1.000	1.000
	F^1 measure	**1.000**	**1.000**

Table 3 Comparison of computation time of the proposed model with other methods

Shot detection method	Per frame feature extraction time (s)
Proposed Model	**0.0032**
Pixel difference [2]	0.0050
ECR [21]	0.1978
Chromaticity histogram [14]	0.0524

The feature extraction time taken by the proposed model and other shot detection models for a single frame are given in Table 3. It is observed from the Tables 1 and 2 that the proposed model is able to detect both hard cuts and fades. The superiority of the proposed model lies in identifying both types of shot boundaries with less computing time and produced accurate shot boundaries, and hence is suitable for video indexing and retrieval applications.

5 Conclusion

We proposed an accurate and computationally efficient model for video cut and fade detection based on cumulative color histogram. Since the dimension of the feature vector is very less, the matching time can be reduced. The proposed model produced better results for both cut and fade detection. The experimental results on TRECVID video database show that the proposed model can be used for real time video shot detection purpose.

References

1. Alattar A (1997) Detecting fade regions in uncompressed video sequences. In: IEEE international conference on acoustics, speech, and signal processing 1999 (ICASSP-97), vol 4. pp 3025–3028

2. Boreczky J, Rowe L (1996) Comparison of video shot boundary detection techniques. J Electron Imaging 5(2):122–128
3. Del Bimbo A (1999) Visual information retrieval. Morgan Kaufmann Publishers Inc., San Francisco
4. Fernando W, Canagarajah C, Bull D (1999) Automatic detection of fade-in and fade-out in video sequences. In: Proceedings of 1999 IEEE international symposium on circuits and systems (ISCAS'99), vol 4, pp 255–258
5. Fernando W, Canagararajah C, Bull D (2000) Fade-in and fade-out detection in video sequences using histograms. In: Proceedings of 2000 IEEE international symposium on circuits and systems (ISCAS 2000), Geneva, vol 4, pp 709–712
6. Guimarães SJF, Couprie M, Araújo AdA, Leite NJ (2003) Video segmentation based on 2d image analysis. Pattern Recogn Lett 24:947–957
7. Hanjalic A (2002) Shot-boundary detection: unraveled and resolved? IEEE Trans Circuits Syst Video Technol 12(2):90–105
8. Kong W, Ding X, Lu H, Ma S (1999) Improvement of shot detection using illumination invariant metric and dynamic threshold selection. In: Visual information and information systems. Springer, Berlin, pp 658–659
9. Koprinska I, Carrato S (2001) Temporal video segmentation: a survey. Signal Process: Image Commun 16(5):477–500
10. Le DD, Satoh S, Ngo TD, Duong DA (2008) A text segmentation based approach to video shot boundary detection. In: 2008 IEEE 10th workshop on multimedia signal processing, pp 702–706
11. Li S, Lee MC (2005) An improved sliding window method for shot change detection. In: SIP'05, pp 464–468
12. Lian S (2011) Automatic video temporal segmentation based on multiple features. Soft Comput—A Fusion Found Methodol Appl 15:469–482
13. Priya GGL, Dominic S (2010) Video cut detection using dominant color features. In: Proceedings of the first international conference on intelligent interactive technologies and multimedia (IITM '10). ACM, New York, pp 130–134
14. Shekar BH, Raghurama Holla K, Sharmila Kumari M (2011) Video cut detection using chromaticity histogram. Inter J Mach Intell 4(3):371–375
15. Smeaton AF, Over P, Doherty AR (2010) Video shot boundary detection: seven years of TRECVID activity. Comput Vis Image Understand 114(4):411–418
16. Truong B, Dorai C, Venkatesh S (2000) Improved fade and dissolve detection for reliable video segmentation. In: Proceedings of 2000 IEEE international conference on image processing, vol 3, pp 961–964
17. Tudor B (2009) Novel automatic video cut detection technique using Gabor filtering. Comput Electr Eng 35(5):712–721
18. Yoo H, Ryoo H, Jang D (2006) Gradual shot boundary detection using localized edge blocks. Multimedia Tools Appl 28(3):283–300
19. Yuan J, Wang H, Xiao L, Zheng W, Li J, Lin F, Zhang B (2007) A formal study of shot boundary detection. IEEE Trans Circuits Syst Video Technol 17(2):168–186
20. Yusoff Y, Christmas WJ, Kittler J (2000) Video shot cut detection using adaptive thresholding. In: Proceedings of the British machine vision conference 2000 (BMVC 2000), Bristol, 11–14 Sept 2000
21. Zabih R, Miller J, Mai K (1999) A feature-based algorithm for detecting and classifying production effects. Multimedia Syst 7:119–128
22. Zhang H, Kankanhalli A, Smoliar SW (1993) Automatic partitioning of full-motion video. Multimedia Syst 1:10–28

Cooperative MIMO and Hop Length Optimization for Cluster Oriented Wireless Sensor Networks

Yuyang Peng, Youngshin Ahn and Jaeho Choi

Abstract An energy-efficient multi-hop communications scheme based on cooperative multiple-input multiple-output technique is proposed for wireless sensor networks. The proposed method takes into account the modulation constellation size and transmission distance and effectively saves energy by optimizing hop distances with respect to the modulation constellation. Here, the equidistance hop length scheme as a conventional scheme is compared to the proposed method. In order to evaluate the performance of the proposed scheme, a qualitative analysis is taken in terms of energy efficiency. The analytical results indicate that the propose method can significantly save total energy consumption in comparison to the conventional one. The possible application of the proposed scheme can be an emergency monitoring system such as the forest fire detection; and this method can provide a longer lifetime of monitoring system.

Keywords Cooperative Communications · Sensor Networks · Multi Hop · Hop Distance · MIMO · Energy Efficiency

Y. Peng · Y. Ahn · J. Choi (✉)
Department of Electronic Engineering, CAIIT Chonbuk National University,
561-756 Chonju, Chonbuk, Republic of Korea
e-mail: yuyang@jbnu.ac.kr

Y. Ahn
e-mail: wave@jbnu.ac.kr

Mohan S. and S. S. Kumar (eds.), *Proceedings of the Fourth International Conference on Signal and Image Processing 2012 (ICSIP 2012)*, Lecture Notes in Electrical Engineering 222,
DOI: 10.1007/978-81-322-1000-9_35,

1 Introduction

It is well known that saving energy is the main concern in wireless sensor networks since the size of the sensors is too small for them to have been designed with enough energy for long time operation. Recently, many approaches including multiple-input multiple-output (MIMO) communication and multi-hop schemes have been developed to improve the energy efficiency. In fact, it is difficult to implement multiple antennas at a small sensor node in a realistic environment for the MIMO approach. Therefore, cooperative MIMO schemes have been designed [1, 2], which allow single antenna nodes to achieve MIMO capability. Energy efficiency has been done to explain that the cooperative MIMO outperforms the SISO (single-input single-output) after a certain distance [3]. But the use of cooperative MIMO in a multi-hop network making the transmission more efficient is still a hot reach topic that should be sought.

In this paper, in order to take full advantage of these approaches, a new scheme is proposed which involves the joint utilization of cooperative MIMO and multi-hop techniques. Moreover, the modulation constellation size is also considered in this scheme for a more practical case. Also, it is further demonstrated that the energy efficiency performance of the proposed scheme can be improved.

The remainder of the paper is organized as follows. In Sect. 2, the fundamentals and system model are introduced. In Sect. 3, the multi-hop cooperative MIMO wireless sensor network is presented; and the quantitative analysis on the performance of the proposed scheme is presented in Sect. 4. Finally, in Sect. 5, the paper is concluded with a brief summary.

2 Fundamentals and System Model

A wireless sensor network composed of n clusters and a destination is shown in Fig. 1b where the cooperative MIMO communication technique Fig. 1a is applied to a multi-hop scheme for saving energy. In the cluster, the longest distance amongst the nodes is defined as d_{long}. The long-haul distance between the nearest nodes of different cluster is defined as d_i ($i = 1, 2, \ldots n$) which is assumed much larger than d_{long}.

In order to evaluate the performance of the proposed scheme, the energy consumption is first discussed. From [1] it is known that the total average power consumption can be categorized into two main components: the power consumption of all the power amplifiers P_{PA} and the power consumption of all the circuit blocks P_c. Considering the optimized transmission time T_{on}, the total energy consumption per bit can defined as follows:

$$E_{\mathrm{bt}} = (1+\alpha)\,\bar{E}_b \times \frac{(4\pi d)^2}{G_t G_r \lambda^2} M_l N_f + P_c T_{on}/L \tag{1}$$

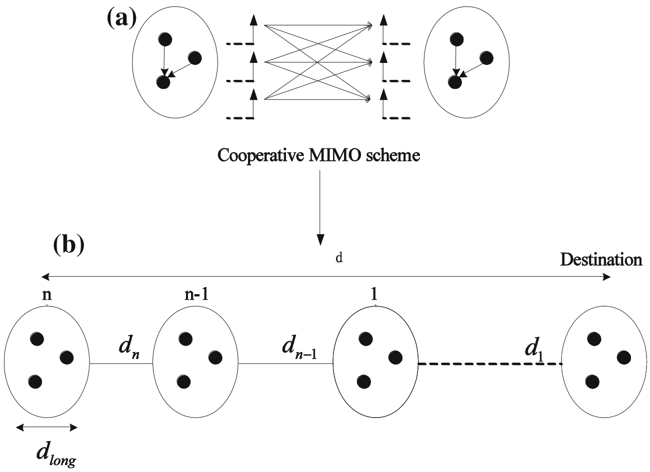


Fig. 1 An energy efficient wireless sensor network is constructed with **a** cooperative MIMO communications and **b** multi-hop techniques

where $\alpha = {}^{\xi}/_{\eta} - 1$ with ξ is the peak to average ratio (PAR); η is the drain efficiency of the RF power amplifiers; $\bar{E}_b$ is the average energy per bit required for a given bit error rate (BER); d is the transmission distance; G_t and G_r are the transmitter and receiver antenna gains, respectively; λ is the carrier wavelength; M_l is the link margin compensating the hardware process variations and other additive interference or background noise; and N_f is the receiver noise. It should be noted that N_f is given by $N_f = N_r/N_0$, where N_r is the power spectral density (PSD) of the total effective noise at the receiver input and N_0 is the single-sided thermal noise PSD at room temperature with $N_0 = -171$ dBm/Hz. Assume that the transmitter buffer length is L bits. In Eq. 1, $\bar{E}_b$ is defined by the BER and constellation size b. The average BER can be obtained as follows [4]:

$$\bar{P}_b \approx \varepsilon_H\left(\frac{4}{b}\left(1 - \frac{1}{2^{\frac{b}{2}}}\right)Q\left(\sqrt{\frac{3b}{M-1}\gamma_b}\right)\right) \tag{2}$$

for $b \geq 2$; and for $b = 1$, Eq. 1 can be simplified as follows:

$$\bar{P}_b \approx \varepsilon_H\left[Q(\sqrt{2\gamma_b})\right] \tag{3}$$

where $\varepsilon_H[\,]$ denotes the expectation given the channel $\mathbf{H}$; γ_b is the instantaneous received signal-to-noise ratio (SNR); $Q(\,)$ is the Q-function. The Chernoff bound can be applied to obtain and the upper bound for $\bar{E}_b$ as follows:

$$\bar{E}_{\mathrm{b}} \leq \frac{2}{3} \times \left(\frac{\bar{P}_b}{\frac{4}{b}\left(1 - \frac{1}{2^{\frac{b}{2}}}\right)} \right)^{-\frac{1}{M_t M_r}} \frac{2^b - 1}{b} M_t N_0 \tag{4}$$

By substituting Eq. 4 into Eq. 1 the total energy consumption per bit can rewritten as follows:

$$E_{\mathrm{bt}} = (1+\alpha) \times \frac{2}{3} \times \left(\frac{\bar{P}_b}{\frac{4}{b}\left(1 - \frac{1}{2^{\frac{b}{2}}}\right)} \right)^{-\frac{1}{M_t M_r}} \times \frac{2^b - 1}{b} M_t N_0 \times \frac{(4\pi d)^2}{G_t G_r \lambda^2} M_l N_f + \frac{P_c}{bB} \tag{5}$$

where B is the modulation bandwidth; M_t and M_r are the number of transmitter and receiver antennas, respectively.

3 Multi-Hop Network Analysis and Calculation

In this section the cooperative MIMO and multi-hop schemes is considered jointly as shown in Fig. 1. Let d_i represent the optimal transmission distance and b_i represent the optimal constellation size. Then, for a transmission distance d_i, the energy consumption per bit of long-haul can be defined as follows:

$$E_{\mathrm{bt}}(d_i) = (1+\alpha) \times \frac{2}{3} \left(\frac{\bar{p}_b}{\frac{4}{b_i}\left(1 - \frac{1}{2^{\frac{b_i}{2}}}\right)} \right)^{-\frac{1}{M_t M_r}} \times \frac{2^{b_i} - 1}{b_i} M_t N_0 \times \frac{(4\pi d_i)^2}{G_t G_r \lambda^2} M_l N_f + \frac{P_c}{b_i B} \tag{6}$$

There can be various scenarios, however, in this paper, it is assumed that two sensor nodes group together making a cluster since this assumption on the number of nodes in a cluster can reduce the complexity of the calculation. Nevertheless, it is important to note that the calculation can in principle be extended to any cluster size. For a scenario where all nodes are transmitting, the total energy consumption can be defined as follows:

$$E_{total} = E_{local} + 2\sum_{i=1}^{n} (n+1-i) E_{\mathrm{bt}}(d_i + 2d_{long}) N_i \tag{7}$$

where each sensor node assumes to transmit N_i bits. E_{local} is the local energy consumption; and $2\sum_{i=1}^{n} (n+1-i) E_{\mathrm{bt}}(d_i + 2d_{long}) N_i$ is defined as the long-haul energy consumption. Using the cooperative communication scheme proposed in

[1, 2] the local energy consumption of the proposed scheme E_{local} can be defined as follows:

$$E_{local} = \sum_{i=1}^{n} \left[\sum_{i=1}^{M_t} N_i E_i^t + \sum_{j=1}^{M_r - 1} N_s n_r E_j^r \right] \tag{8}$$

where E_i^t, E_j^r denotes the energy cost per bit for local transmission for the transmitter and receiver, respectively; N_s is the total number of symbols; n_r is the number of bits after quantizing a symbol at the receiver.

On the other hand, in order to minimize the total energy consumption, in this paper, the proposed method tries to optimize the hop distance between the clusters. Let d_i be the optimal transmission distance. First, through observation of the transmission distance in the proposed model, the constraint $\sum_{i=1}^{n} d_i = d - nd_{long}$ is set. Along with the constraint, the cost function can be defined as follows:

$$\Phi = E_{local} + 2 \sum_{i=1}^{n} (n + 1 - i) E_{\mathrm{bt}} (d_i + 2d_{long}) N_i + w(d - nd_{long} - \sum_{i=1}^{n} d_i). \tag{9}$$

To Minimize E_{total} under the constraint $\sum_{i=1}^{n} d_i = d - nd_{long}$, the partial derivatives with respect to d_i are taken and set them equal to 0 as follows:

$$\frac{\partial \Phi}{\partial d_i} = 4E \times (n + 1 - i)(d_i + 2d_{long}) - w = 0 \tag{10}$$

where E is $(1+\alpha) \times \frac{2}{3} \times \left(\frac{\bar{p_b}}{\frac{4}{b_i}\left(1 - \frac{1}{2^{\frac{b_i}{2}}}\right)} \right)^{-\frac{1}{M_t M_r}} \times \frac{2^{b_i} - 1}{b_i} M_t N_0 \times \frac{(4\pi)^2}{G_t G_r \lambda^2} M_l N_f N_i$;
and w is a Langrage's multiplier.

Solving for d_i in Eq. 10, the hop distance between the clusters can be obtained as follows:

$$d_i = \left(\frac{w}{4E(n + 1 - i)} \right) - 2d_{long} \tag{11}$$

where w can be obtained by using the constraint $\sum_{i=1}^{n} d_i = d - nd_{long}$. Then Eq. 11 can be rewritten as follows:

$$d_i = \frac{d + nd_{long}}{\sum_{f=1}^{n} (\frac{1}{n+1-f})(n + 1 - i)} - 2d_{long}. \tag{12}$$

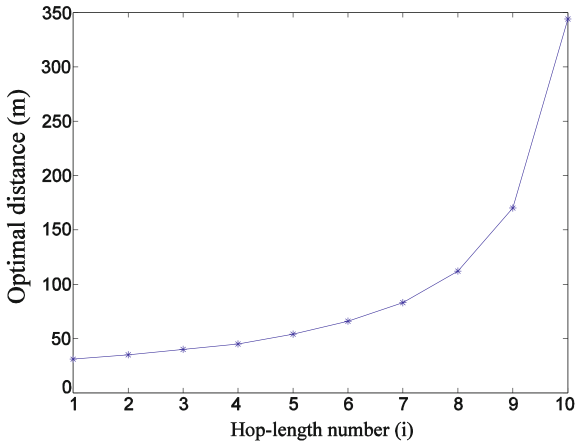


Fig. 2 Optimal distances versus hop-length numbers

4 Numerical Results

In this section, a quantitative analysis on the performance of the proposed method is presented. There are various scenarios to be considered, however, for the preliminary study, some to the parameters are set in such a way to simplify the complexity of the equations and also the computation.

Now, suppose that $n = 10$, $d = 1000$ m, and $d_{long} = 2$ m. In Fig. 2, the optimal transmission distances are obtained using Eq. 12 and the results are plotted. The figure shows that for the clusters located farther from the destination the hop distances or lengths for them also increase. The reduction of energy consumption can be obtained using this scheme instead of the equidistance scheme. In order to compare the performance of the proposed method with the equidistance scheme, one set of typical parameters are set and used as follows:

$B = 10$ kHz, $f_c = 2.5$ GHz, $P_{mix} = 30.3$ mW, $P_{filt} = 2.5$ mW, $P_{filr} = 2.5$ mW, $P_{LNA} = 20$ mW, $P_{synth} = 50$ mW, $M_l = 40$ dB, $N_f = 10$ dB, $G_tG_r = 5$ dBi and $\eta = 0.35$, $N_i = 20$ kb, $n_r = 10$.

For simplicity, Alamouti scheme [5] has been adopted in this paper. Using the brute-force simulation method proposed in [6] the optimal constellation size b is obtained for different transmission distances as listed in Table 1.

For the evaluation of the performance, an equidistance scheme is calculated in order to compare with the proposed scheme. As already referred, the proposed scheme is the use of cooperative MIMO in optimal hop-length wireless sensor network with consideration of modulation constellation size, extra training overhead requirement, and data aggregation energy. In Fig. 3 the total energy consumption per bit is plotted as a calculation aim for the proposed scheme and equidistance scheme. It can be seen that the majority of clusters in the proposed scheme have less total energy consumption per bit when compared with the equidistance scheme, i.e. the proposed scheme has a better performance. After calculating energy consumption of each cluster and adding all the values together,

Table 1 Optimal constellation size b versus different transmission distances

d(m)	10	20	40	70	100	150	200	300	350
b	8	6	5	4	3	2	2	1	1

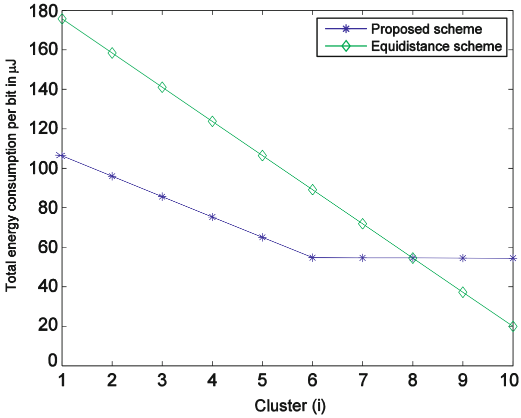


Fig. 3 Total energy consumption per bit in every cluster for different schemes ($n = 10$, $d = 1000$ m, $d_{long} = 2$ m).

the results show that the proposed scheme offers a total energy saving of about 29.5 %.

5 Conclusion

A multi-hop scheme based on a cooperative MIMO has been proposed. The feasibility of using this scheme for optimization of the performance of the wireless sensor networks is validated numerically by measurement of the total energy consumption. The results demonstrate that the proposed scheme offers a total energy saving of around 29.5 % after taking into account with modulation constellation size and transmission distance compared with the traditional scheme. Therefore it is concluded that the proposed scheme can be applied in wireless sensor networks for the reduction of energy consumption when the prime concern is to extend the network life time.

References

1. Cui S, Goldsmith A, Bahai A (2004) Energy-efficiency of MIMO and cooperative MIMO techniques in sensor network. IEEE J Select Areas Comm 22:1089–1098
2. Gai Y, Zhang L, Shan X (2007) Energy efficiency of cooperative MIMO with data aggregation in wireless sensor networks. Proceedings of IEEE Wireless Comm. Networking Conference, pp 791–796

3. Cui S, Goldsmith A, Bahai A (2003) Modulation optimization under energy constraints. Proceedings of IEEE international conference on comm, pp 2805–2811
4. Proakis J (2000) Digital Communications, 4th edn. McGraw Hill, New York
5. Alamouti S (1998) A simple transmit diversity technique for wireless communications. IEEE Jr Select Areas Comm 16:1451–1458
6. Cui S, Goldsmith A, Bahai A (2005) Energy-constrained modulation optimization. IEEE Trans Wireless Comm 4:2349–2360

Performance Analysis of Feature Point Detectors in SFF-Inspired Relative Depth Estimation

R. Senthilnathan and R. Sivaramakrishnan

Abstract This paper is a part of the research that attempts to develop a new method for estimating relative depth of scene based on focus cue inspired by the Shape From Focus (SFF) technique which basically is a 3-D vision technique aiming at scene reconstruction. The proposed method is developed for scene with objects whose geometrical dimensions are of macro-level unlike the conventional SFF algorithms which could deal with extremely small objects in order to avoid parallax. This is essentially because SFF techniques involve motion of either the camera or the object causing structure dependent pixel motion in the image. This hinders finding the corresponding pixels in the sequence of images to perform a measure of image focus and resulting in erroneous reconstruction. The parallax effect is tackled using a matching technique which tracks point correspondences in the image sequence. The points with good local contrast are extracted using the so-called interest point detectors. The paper analyses five different point detectors used to extract feature points in the image. The point detectors were analysed for a number of parameters like repeatability, false detects, matchability and information content. The paper compares their performance and attempts to bring out their advantages and limitations. Three different conditions viz., uniform illumination, non-linear illumination changes, presence of impulse noise in the image were considered. The image processing method is validated for two different textures, one being a repetitive pattern and other a non-repetitive pattern.

Keywords Shape from focus · Parallax · Image blur · Sparse reconstruction · Feature point detector · Quantitative evaluation

R. Senthilnathan (✉) · R. Sivaramakrishnan
Division of Mechatronics, Department of Production Technology, MIT Campus, Anna University, Chennai, India
e-mail: sen2dext@gmail.com

Mohan S. and S. S. Kumar (eds.), *Proceedings of the Fourth International Conference on Signal and Image Processing 2012 (ICSIP 2012)*, Lecture Notes in Electrical Engineering 222, DOI: 10.1007/978-81-322-1000-9_36,

1 Introduction

Computer vision has been increasingly finding its application in many fields like automated inspection, robot guidance, entertainment and other scientific and industrial applications. Scene reconstruction in the field of 3-D Vision has been the topic of interest for over three decades. Many techniques viz. passive, active and hybrid techniques have evolved over the years. Passive methods for 3-D reconstruction aim at estimating the 3-D surface geometry with one or more passive cameras which records intensity information of the scene, while active vision techniques reconstructs a scene by estimating the depth by deliberately releasing some form of energy into the scene which was unless not present in the scene. Methods such as the structured light technique combine the advantages of active and passive vision techniques by throwing additional light patterns on the scene which is imaged by a passive camera. The techniques are generally grouped together under the name Shape-from-X. Where X denotes the cue used for the scene reconstruction which could be stereo, motion, shading, focus, defocus, texture etc. Out of these methods Shape from Focus (SFF) [1] and Shape from Defocus (SFD) use multiple images of the scene taken with different focus. The difference between the methods come the in the form that the SFD [2] generally require one sharp image of the foreground and one sharp image of the background. The distance of all the points that lie between foreground and background is interpolated by a sharpness measure (actually measures the degree of un-sharpness). Unlike the SFD method, SFF generally requires more number of images along the different focal stacks of the lens. This generally makes the SFF method computationally expensive compared to SFD. The SFF techniques demand images of the scene with different focus levels which can be obtained by translating the camera or the object or by changing the focus setting of the lens. One of the inherent limitations of the SFF method of reconstruction of the scene is that they are highly sensitive to parallax. The paper is out of this centre theme of the research work dealing with the parallax in a SFF-inspired algorithm.

2 Camera Modelling

Modelling a camera could mean a variety of modelling ranging from physical models to geometrical representations. Generally for tasks like those considered in the paper, the geometrical perspective of image formation is relevant compared to other perspectives. The paper opts to model the camera used in the work as a pin-hole camera model. A simple camera model consisting of thin lens and an image plane can be used to derive some fundamental characteristics of image formation and focusing. Figure 1 shows the basic geometry of image formation of focused and defocused scenes. Suppose the object is stationary and the images are obtained on image plane by translating the lens (camera) along the optical axis. All light rays,

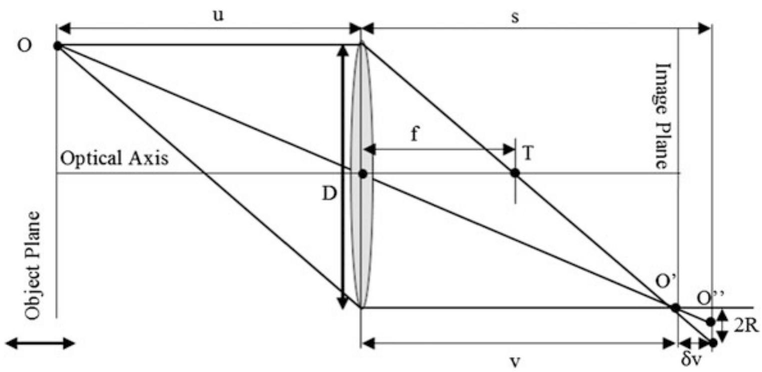


Fig. 1 Image formation model

which are radiated from the object, are intercepted by the lens and converged at the image plane. A point O on the object is well focused and its image is obtained at point O' on image detector. A well-focused point O satisfies the lens law:

$$\frac{1}{f} = \frac{1}{u} + \frac{1}{v} \tag{1}$$

The images acquired from projection needs to reach the image plane at the focal point so that points in reality are mapped to the points in the image Points which are not in focus appear blurred. If point O is not in focus then it gives rise to a circular image called the blur circle on the image plane (the image sensor). According to geometric optics, the intensity within the blur circle is approximately constant.

But due to aberrations, diffractions and other undesirable effects a two-dimensional Gaussian model has been suggested as an alternative model:

$$h(x, y) = \frac{1}{2\pi\sigma^2} e^{\frac{-1}{2}\left(\frac{x^2+y^2}{\sigma^2}\right)} \tag{2}$$

where is a spread parameter such that $\sigma = \mathrm{k.\ d}$ for $\mathrm{k} > 0$. k is a proportionality constant which is a characteristic of the camera used for imaging. The value of k can be found from calibrating the camera. This blur model is often referred to as Point Spread Function (PSF).

3 Shape from Focus

The input in the conventional SFF technique can be regarded as a sequence of images that correspond to different levels of focus of the objects in the scene. In this case the images are obtained by moving the camera towards the stationary object. The basic idea in SFF techniques is to find the frame from the sequences of frames where the different depths of objects are in focus or nearly focused. In other words for every pixel the frame from the input sequence of images in which the corresponding pixel exhibits the maximum sharpness has to be selected. The best focus or maximum sharpness can be found out for every pixel by using a focus measure. In this case sum of modified Laplacian is used as the focus measure given by the following expression.

$$\mathrm{SML}(x_0, y_0) = \sum\nolimits_{p(x,y) \in \cup(x_0,y_0)} \left(\frac{\partial^2 g(x,y)}{\partial x^2}\right)^2 + \left(\frac{\partial^2 g(x,y)}{\partial y^2}\right)^2 \tag{3}$$

where g (x, y) is the input image frame and p (x, y) is a pixel in the neighbourhood U (x_0, y_0) of pixel (x_0, y_0). And the other important issue in SFF techniques is that since only a discrete number of frames are used there is definite chance of loss of information between consecutive frames. In the work carried out, a fixed focal length wide angle lens is used hence the probability of information loss is high due to parallax. This leads to inaccurate results for the depth map. Hence to tackle this issue the corresponding points in the image are tracked in the sequence so as to minimize the error. Only points with good features could be tracked and it's for this reason the feature point detectors are required. It can be shown that identically defocused image out of the sequence of images acquired does not exist. It can also be shown that changing the image fine focus setting of the lens (image distance), aperture or the focal length itself to obtain a blur equivalent to the one obtained by translating is not possible since all the changes are also function of the depth of the points of consideration in the scene. This is simply because in each case they become a function of depth of scene.

4 Experimental Setup

To ensure pixel motion that could be caused by translation along the optical axis, a precise mechanism is required. The experimental translating mechanism is based on a surf coder mechanism consisting of a worm and pinion gear coupled to a timing belt drive. The photograph of the experimental setup along with the image showing the internal construction is shown in Fig. 2.

The images are acquired by synchronizing the motor speed with camera frame rate with the help of the incremental encoder such that for every 1 mm of motion along the optical axis, one image is acquired. The total travel is restricted is restricted to 30 mm hence 30 frame are acquired to detect. Though anything more than 20 images are sufficient for SFF techniques, always more is better. Some sample images of the two textures considered are shown in Fig. 3.

The detailed specifications of the various elements of the system are listed in Table 1.

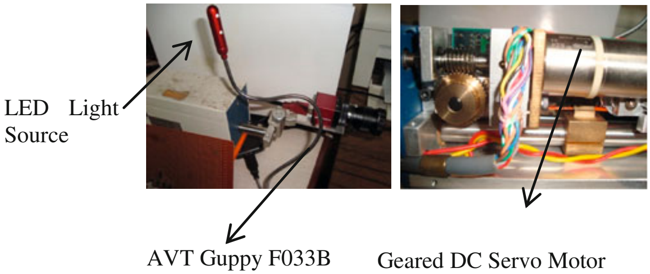


Fig. 2 Experimental setup

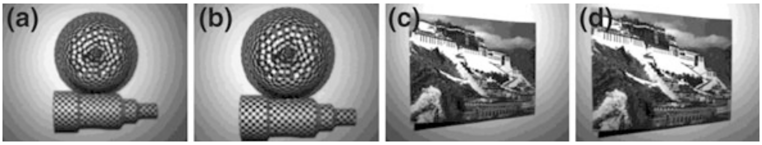


Fig. 3 Sample images under uniform illumination: repetitively textured objects, **a** 1st frame, **b** 30th frame, object with non-repetitive texture, **c** 1st fame, **d** 30th frame

Table 1 System specifications

Specification	Value
Lens focal length	16 mm, C-mount
Aperture, f#	1.3
Working distance	Variable
Camera	AVT guppy F033B,Area scan, Monochrome progressive
Camera interface	IEEE 1,394 a to PCI bus
Sensor size	1/3″ and 4:3 aspect ratio
Sensor resolution	640 × 480
Frame rate	30 fps
Light source	White LED
Lighting technique	Partially diffused bright field incident lighting
Scene illuminance	160 lux (uniform lighting case)
Hardware platform	2.6 GHz Quad Core with 2 GB primary memory and 1,300 MHz bus speed
Software platform	Microsoft visual studio (VC++, Halcon, MATLAB, IPT)
Image acquisition time	180 ms per image

5 Interest Points

As mentioned earlier the paper is an attempt to have a measurement of sparse relative depth of the scene. The parallax problem is tackled by observing a set of significant points over the entire sequence of images. Tracking for the same points on the scene is possible only if they have good contrast in the local neighbourhood.

The feature points are those points where the image function changes in two dimensions. The changes could be either observed as junction points where the image edges intersect or as area points where color or brightness differs from the surroundings. If we could extract these information which are basically affected by focus of the lens, a measure of the depth is possible. Having said this it is very clear that extracting these points with good accuracy are extremely important. The following section if the paper initially explores the mathematics of feature point extraction for five different approaches.

5.1 Forstner Corner Detector

The first method is based on the famous method for detecting feature points presented by Forstner and Gulch in [3]. The method generally involves a two-step procedure. In the first step point regions are extracted by searching for optimal windows and the second step involves locating optimal points within the selected window by calculating two optimization functions for the resulting points in the first step. The image is initially smoothened a matrix operation given in the following equation in which $I_{x,\ c}$ and $I_{y,\ c}$ are the first derivatives of each image channel i.e., along column and row dimensions of the image. S is the smoothing function basically incorporated to reduce noise by means of average smoothing.

$$M = S^* \begin{pmatrix} \sum_{c=1}^{n} I_{x,c}^2 & \sum_{c=1}^{n} I_{x,c} I_{y,c} \\ \sum_{c=1}^{n} I_{x,c} I_{y,c} & \sum_{c=1}^{n} I_{y,c}^2 \end{pmatrix} \tag{4}$$

This is required so as eliminate any random noise in the image in the form of local gray level deviations which does not correspond to real objects colour or brightness. Once the smoothening is complete, as a second step the two eigenvalues λ_1 and λ_2 of the inverse of M are taken as interest values into account. The eigenvalues basically define the axis of an error ellipse. In our implementation, the operator computes inhomogeneity and isotropy of the texture in the resultant image M. These measures are required for feature point detection operator after all a point in the image to be called a good feature point, should have clear distinction from its neighbourhood.

5.2 Harris Corner Detector

The Harris corner detector algorithm was proposed by C. Harris and M. Stephens in [4]. The Harris operator not only solves the problem of discrete pixel shifts, but also deals with the direction issue thus increasing the accuracy of localization. If the similarity comparison for each point-shift changes, it is an indication of presence of a feature point. Similar to the Forstner operator, the Harris method also is based upon the smoothened matrix M with respect to Eq. 4. The smoothening is performed by a Gaussian with $\sigma = 2$ and the Gaussian derivatives as mentioned in the above equation are of size 1. The Harris operator basically determines the point weight R from the matrix M with corner response function:

$$R = Det(M) - \alpha \cdot (Trace(M))^2 \tag{5}$$

The parameter α is empirically selected between 0.03–0.06 to differentiate points and edges. This yields positive values for points and negative values for straight edges. The positive local extrema of R represents the feature points.

5.3 Harris Corner Detector with Binomial Approximation

The paper presents one more operator which deviates itself from the traditional Harris operator presented in the last section by using a binomial smoothing instead of a Gaussian. In fact binomial filter is a very good approximation of a Gaussian which can be implemented using integer operations alone. Let m be the mask height and n, the mask width, the filter coefficients b_{ij} are given in the following equation, where i = 0, 1,..., $m-1$ and $j = 0, 1,\ldots, n-1$. The smoothing function is applied to the image derivative using a Sobel mask of size 5 × 5.

$$b_{ij} = \frac{1}{2^{n+m-2}} \binom{m-1}{i} \binom{n-1}{j} \tag{6}$$

Once smoothened, the point weight R is computed using the Eq. 5 as in the previous case with a Gaussian smoothing. In the same lines the interest points are located from the positive local extrema of R.

5.4 Sojka Corner Detector

The Sojka corner detector presented by Sojka in [5] computes the corner response function combining the angle and the contrast of the corners in the image. The function is designed in such as way that it exhibits its local maxima at corner points. The implementation of the Sojka corner detector defines corner as a point where two straight, non-collinear gray value edges intersect. To identify a corner a neighbourhood of size 9 × 9 is inspected for relevancy. Pixels with a gradient whose magnitude is less than 30 are ignored from the outset. The corner response function of the Sojka operator implementation for a point in the image say P with a neighbourhood Ω is given by

$$R = g(P)\sigma_{\varphi}^2(P) \tag{7}$$

where $g(P)$ is the magnitude of gradient brightness $\sigma_{\varphi}^{2}(P)$ is the average weighted square value of the difference between the edge direction and the average edge direction.

5.5 Lepetit Corner Detector

The Lepetit corner detector implementation is based on the method presented by Lepetit et al. in [6]. The image is first smoothed with a median of size 3×3. A circle with radius of 3 pixels is chosen around the interest point and examined for being a candidate point of good feature. Then the absolute differences of two diagonally opposed gray values say m_1 and m_2 on the circle to the centre of the circle m is computed. A threshold of 15 is set and at least one of the differences computed has to be larger than the threshold. Those pixels which satisfy this condition are taken as interest points.

6 Point Correspondences

The feature points are tracked across the sequence of images so that focus can be measured in all the corresponding points. Tracking of points is subjected to existence of a point in all the frames. An operator's performance is hindered by a number of factors such as pixel motion, occlusion, image blur, changes in lighting conditions etc. Out of the mentioned factors it is very obvious that our system inherently involves changes in focus causing blur in image. Hence before matching of points to find their corresponding location in two consecutive frames, their existence must first be ensured. This is taken care by the matching step. Once the characteristic points are determined in any two consecutive images, the corresponding locations of all the points in both images should be determined. It can be shown that any two consecutive images in the sequence are geometrically related by a well-defined projective transformation matrix, generally known as the fundamental matrix. The matching process is carried out in two steps. Initially correlations of the gray values around the input points in the first and the second image are determined and an initial matching between them is generated using the similarity of the windows in the two images. The size of the mask window is taken as 5×5. A small window is taken since interest points have a good local difference in gray values. Large masks would tend to find erroneous correspondences. The correlation procedure adopted is a *normalized cross correlation* (NCC). Once the initial matching is complete, a randomized search algorithm namely *Random Sample Consensus* (RANSAC) algorithm is applied [7]. The algorithm basically estimates a more accurate fundamental matrix that allows finding the final correspondences with greater precision. This matrix is further optimized using all consistent points using the gold standard method. Since the analysis of matching technique is not within the scope of the paper, the algorithms are not elaborated further.

7 Evaluation Criteria

The research work is completely based on point features since the aim is only to obtain an estimate of relative depth of the scene. The point detection algorithms presented in the previous section though are proven methods, have their own unique advantages and disadvantages. Based on the demand from the application of point detectors for the case considered in the paper the following criteria were found to be more relevant to have an idea of the merits and demerits of the operators. *Repeatability* of detecting the same point in set of images in the sequence acquired under conditions. The definition of the parameters is based on the ones considered in [8] but the parameters are evaluated for 30 frames contradicting with [8] where the parameters are evaluated only for two images. Moreover explicit application of such point operator evaluation for SFF systems is completely new since SFF systems usually aims at dense scene features rather than point information. The other major deviation from the conventional analysis procedure is the dynamic selection of reference image. Generally when parameters like repeatability rate are evaluated a reference is manually selected where in this case any frame in the sequence can be a reference image selected based on a predefined criterion based on an indirect evaluation of image focus on the whole. In this paper the following conditions are considered.

- In the parallax effect which can be modelled as change in viewpoint (as the camera is moves towards the object) resulting in change of scaling (U.L.).
- With respect to illumination changes and shadows (N.U.L).
- In the presence of synthetically generated white noise of amplitude 30 (W.N.).

7.1 Repeatability

Once the point correspondences in the sequence are found repeatability rate RR_i in frame i can be defined as:

$$RR_i = \frac{NP_i}{NP_r} \tag{8}$$

where NP_i is the number of interest points detected in frame i and NP_r is the number of interest points detected in the reference frame, r. As mentioned in last section reference image is that image in the sequence which has the highest contrast and best focus. A high contrast image by principle is where point detection is easy compared to a poor contrast image. From the definition it could be realized that $RR_i = 1$ indicates that all the points in the reference frame are detected in the frame i, which is an essentially characteristic of a perfect detector.

8 Results and Discussion

The results of the repeatability rate for the various operators for three different conditions are tabulated in Table 2.

First, observations on the time performance the Harris operator with the binomial smoothing was the fastest of all while Sojka implementation took the longest time for detecting feature points. Under uniform lighting conditions detection was predictable for all the algorithms. Forstner's method returned the least number of points nut was very robust with least number of false detects where as Sojka and Lepetit operators returned more points and more false detections. The trade-off could be based on the application in general though in our application lesser the number of false detections better is the estimate of relative depth. The performance of the operators with random illumination changes introduced by casting dynamic shadows on the object was interesting. All operators suffered from poor repeatability rate and less number of points. It's essentially because of loss of sufficient contrast in the local neighbourhood. The Forstner operator proved to be better with a higher repeatability rate. Introduction noise on the images of

Table 2 Experimental results

Operator		Repetitive texture		Non-repetitive texture		Average processing time for an image (s)
		Average no. of points	Average repeatability rate	Average no. of points	Average repeatability rate	
Forstner	U.L.	396	0.9662	425	0.9612	0.1146
	N.U.L.	407	0.9240	473	0.8962	
	W.N.	602	0.9397	816	0.9403	
Harris (G)	U.L.	551	0.9665	925	0.9251	0.1087
	N.U.L.	580	0.8671	957	0.8957	
	W.N.	2,410	0.9514	2,760	0.9825	
Harris (B)	U.L.	680	0.9748	924	0.9362	0.0935
	N.U.L.	700	0.9150	990	0.8893	
	W.N.	1,056	0.9659	1,702	0.9316	
Lepetit	U.L.	1,688	0.9202	1,459	0.9663	0.0939
	N.U.L.	1,612	0.9020	1,846	0.8612	
	W.N.	2,179	0.8902	1,814	0.9567	
Sojka	U.L.	996	0.9282	1,726	0.9833	0.1593
	N.U.L.	1,004	0.9114	2,171	0.8802	
	W.N.	2,040	0.9406	3,178	0.9579	

uniformly illuminated scene corrupted the contrast and introduced false contrasts which were detected as interest points by Sojka and Lepetit implementation of our work. This is possibly because of the algorithms ability and parameters used for image smoothing. Tuning them to optimal numbers of course would yield better results.

References

1. Nayar SK, Nakagawa Y (1990) Shape from Focus: an Effective approach for rough surfaces CRA90. pp 218–225
2. Xiong Y, Shafer SA (1993) Depth from focusing and defocusing. IEEE Comput Vis Pattern Recogn 68–73
3. Förstner W, Gülch E (1987) A fast operator for detection and precise location of distinct points, corners and circular features. In: Proceedings of the intercommission conference on fast processing of photogrammetric data. pp 281–305
4. Harris C, Stephens M (1988) A combined corner and edge detector. In: Proceedings of 4th Alvey vision conference. Manchester, 147–151
5. Sojka E (2002) A new algorithm for detecting corners in digital images. In: Proceedings of SCCG 2002. ACM SIGGRAPH, NY, pp 55–62
6. Lepetit V, Fua P (2006) Key point recognition using randomized trees. IEEE Trans Pattern Anal Mach Intell 28(9):1465–1479
7. Fischler MA, Bolles RC (1981) Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Commun ACM 24(6):381–395
8. Schmid C, Mohr R, Bauckhage C (2000) Evaluation of interest point detectors. Int J Comput Vis 37(2):151–172
9. Helmli FS, Scherer S (2001) Adaptive shape from focus with an error estimation in light microscopy. In: Proceedings of the 2nd international symposium on image and signal processing and analysis
10. Sahay RR, Rajagopalan AN (2011) Dealing With parallax in shape-from-focus. IEEE Trans Image Process 20(2):558–569
11. Moravec H (1980) Obstacle avoidance and navigation in the real world by a seeing robot rover, Tech report CMU-RI-TR-3. Carnegie-Mellon University, Robotics Institute

An Efficient Classification Analysis for Multivariate Coronary Artery Disease Data Patterns Using Distinguished Classifier Techniques

G. NaliniPriya, A. Kannan and P. Anandhakumar

Abstract Medical care industry has huge amount of data, which includes hidden information. Advanced data mining techniques can be used to develop classification models from these techniques for effective decision making. A system for efficient and automated medical diagnosis would increase medical care and reduce costs. This paper intends to provide a survey of current techniques of knowledge discovery in databases using data mining techniques that are very much needed for current study in medical research predominantly in Heart Disease diagnosis. The data mining classification techniques such as K means, SOM, decision Tree Techniques are explored with the algorithm for coronary Artery disease dataset (CAD) taken from University California Irvine (UCI). Performance of these techniques are compared through standard metrics. Number of experiment has been conducted to evaluate the performance of predictive data mining technique on the same dataset. The output shows that Decision Tree outperforms compared to other classifiers.

Keywords Knowledge discovery · Data mining · CAD · Heart disease · Classification · Multivariate data · Decision tree

G. NaliniPriya (✉) · A. Kannan
Department of Information Science and Technology, Anna University, Chennai, India
e-mail: nalini.anbu@gmail.com

P. Anandhakumar
Department of IT, Anna University, Chennai, India

Mohan S. and S. S. Kumar (eds.), *Proceedings of the Fourth International Conference on Signal and Image Processing 2012 (ICSIP 2012)*, Lecture Notes in Electrical Engineering 222,
DOI: 10.1007/978-81-322-1000-9_37,

1 Introduction

Data mining technology renders a user related approach to concealed patterns in the data. The smart hospitals today employ some sort of medical information systems to manage their patient data. These systems typically generate huge amounts of data which take the form of numbers, text, charts and images. Modern hospital practices depend upon modern computer based technologies. Data mining algorithms play major role in designing computing environments for prediction of disease. For example, classification algorithms are often useful in patient activity classification [1] and the diagnosis [2–4] of a disease using a multivariate clinical data, which were acquired from the hospital environment using different technologies. This data may be the combination of different types. Advanced development of computer based prediction methods for the diagnosis of heart disease attracts many researchers. In the past time, the use of computers was to build knowledge [5, 6] based decision support system which uses knowledge from medical experts and transfer this knowledge into computer algorithms manually. This process is time consuming and really depends on medical expert's opinion which may be subjective. New classifiers are developed to help the physician in their diagnosis process. In this paper we classify the Multivariate CAD data using well known algorithms and study the performance to gain knowledge from raw data.

1.1 Related Work

Heart disease, which is usually called coronary artery disease (CAD), is a term that can refer to any condition that affects the function of heart. These papers classifying the heart data using decision tree with different algorithm., Many CAD patients have symptoms like chest pain (angina) and fatigue, normally it occur when the heart isn't receiving adequate oxygen. Most of the [3, 7, 8] patients, however, have no symptoms until a heart attack occurs. A number of factors have been shown to increase the risk of developing CAD. Many medical data also need smart technologies to classify the given data or patterns. These algorithms [5, 6, 9] applied a kernel induced metric to assign each piece of data and experimental results were obtained using a well known benchmark of heart disease [10] datasets.

In our proposed method we make use of UCI dataset. It is a well known dataset [1, 7] for research work. The heart disease database from the UCI KDD Archive was used. Many papers handled this bench mark dataset [10, 11] for research purposes. This is a multivariate data is very complex in nature. Motivated by the need of classifying CAD Multivariate data we are going to construct a classification model with decision tree, SOM and K means classifiers based on the dataset and its attributes. Satisfied result has been occurred by making use of proposed algorithm. An experimental comparison is made between the existing methods and the results are analyzed based on their performance.

1.2 About Coronary Artery Disease Dataset

When people talk about heart disease, they usually mean coronary artery disease (CAD). It is the most common type of heart disease. With CAD, plaque builds up on the walls of the arteries that Heart disease, which is usually called coronary artery disease (CAD), is a broad term that can refer to any condition that affects the heart. CAD is a chronic disease [9, 12] in which the coronary arteries gradually hardens and narrow which carry blood to the heart. Over time, this build-up causes the arteries to narrow and harden, called atherosclerosis. This can lead to Angina chest pain or discomfort that happens when the heart doesn't get enough blood. Heart attack happens when a clot mostly or completely blocks blood flow to the heart muscle. Without blood the heart will start to die. If a person survives a heart attack, the injured area of the heart muscle is replaced by scar tissue. This weakens the pumping action of the heart. So, diagnosing the [8, 13, 14] heart disease of a patient involves highly skilled physicians and experience. Moreover computational tools have been designed to improve the abilities of physicians for making decisions about their patients. In this paper, a new classification Analysis is made for [5, 15] multivariate CAD dataset is proposed for useful storage, retrieval and research.

2 Proposed Analysis Work

2.1 Decision Tree

A Decision Tree is a tree-structured plan which contains a attributes to test in order to forecast the output.

A decision tree based IDTC algorithm [7] for classifying Multivariate heart disease is detailed in Fig. 1. The value of Spn is empirically chosen to grow the decision tree for classification from the set of attributes Ajn. Compare each Iv value from record with ps, where Iv is input value from record. Based on the split value spn the attributes are placed left or right of the decision tree. Here class labels "1" for normal and label "2" for Sick. The algorithm handles two class classification problems with respect to class labels. The Figs. 2 and 3 explains the K Means algorithm and SOM steps to execute the process of classifying Multivariate data.

Input: Coronary Artery Disease Data

Step 1: Assign set of attributes as SA, and point values as ps,.,.
SA={Aj1,Aj2....,Ajn}
Step 2: Select the attribute Ajn from SA as a root node and convert its value into point value(ps)
Step 3: Check for Ci ϵC , where C Set of class labels and calculate point values for each attributes Ajn, where ps={p1,p2...,pn}
Step 4: Assign Input values as Iv, Where Iv is set of input values from the tuples/records.
Step 5: Place the tuples into two subset "left" and "right" based on input value (Iv)
Step 6: Test the condition, If Iv,≤ Sv Completely lies Left of decision tree.
Step 7: Test the condition, If Sv ≤ Iv Completely lies Right of decision tree.
Step 8: Check for pruning level at any point of the node, to avoid testing error.
Step 9: Check for Ci at each leaf node .of the tuples.
Step 10: Repeat the steps 1 to 9 for classify the tuples.

Output: Records are classified according to their Class labels

Fig. 1 Steps of intelligent decision tree construction [IDTC] algorithm

2.2 The K Means Algorithm

Input: Coronary Artery Disease Data

Step 1: Selection of the initial k means for k clusters,
Step 2: Calculation of the dissimilarity between an object and the mean of a cluster,
Step 3: Allocation of an object to the cluster whose mean is nearest to the object,
Step 4: Re-calculation of the mean of a cluster from the objects allocated to it in such a way that the intra cluster dissimilarity is minimized.
Step 5: Except for the first operation, the other three operations are repeatedly performed in the algorithm until the algorithm converges.

Output: Classified Patterns

Fig. 2 K means clustering for multivariate data

Input: Coronary Artery Disease Data set

Step 1: Randomize the map's nodes' weight vectors
Step 2: Grab an input vector Traverse each node in the map
Step 3: Use Euclidean distance formula to find similarity between the input vector and the map's node's weight vector
Step 4: Track the node(BMC) that produces the smallest distance
Step 5: Update the nodes in the neighborhood of BMU by pulling them closer to the input vector $Wv(t + 1) = Wv(t) + \Theta(t)\alpha(t)(D(t) - Wv(t))$

Output: Classified Patterns

Fig. 3 SOM for multivariate data

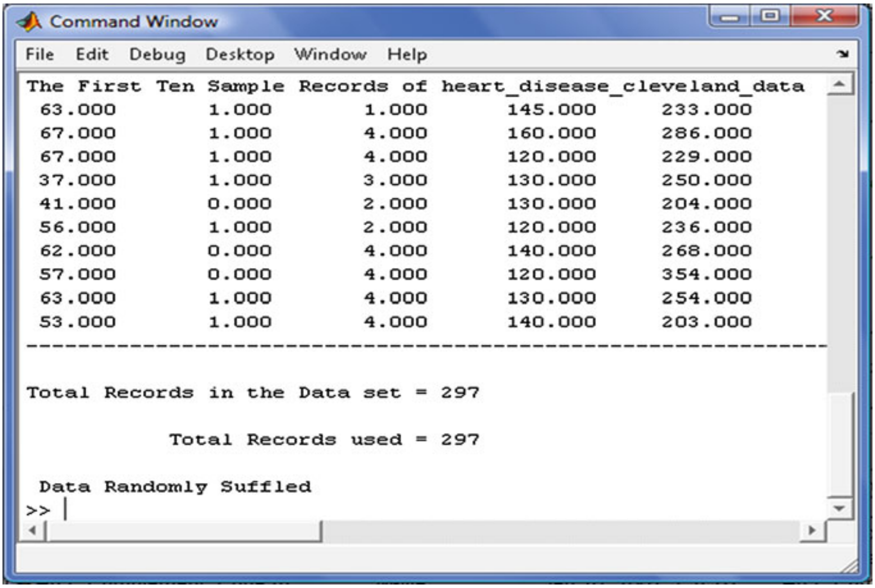


Fig. 4 The screen shots of loading CAD data patterns into MATLAB

2.3 *The SOM (Self-Organizing Map) Algorithm*

There are two ways to interpret a SOM. Because in the training phase weights of the whole neighborhood are moved in the same direction, similar items tend to excite adjacent neurons. Therefore, SOM forms a semantic map where similar samples are mapped close together and dissimilar apart. The other way to perceive the neuronal weights [16, 17] is to think them as pointers to the input space. They form a discrete approximation of the distribution of training samples. More neurons point to regions with high training sample concentration and fewer where the samples are scarce.

3 Implementation Details

To evaluate the algorithms under consideration, a suitable and standard multivariate data set is needed. A suitable [10, 18, 19, 23], UCI data set called “cleveland data” is utilized in this analysis work. This database contains 303 records with 13 attributes which have been originally extracted from a larger set of 75 attributes and a class attributes. Among the 303 records, 164 belong to healthy and remaining are from diseased. The simulation of the Decision tree, means and SOM are constructed by using MATLAB (Release 2009a) Statistical Tool Box. The Fig. 4 represents the sample Medical dataset utilized in Matlab simulation.

The decision tree classification [1, 5, 7, 20] was done using the above algorithms for several times and the best results were tabulated. Here we are taking pruning level is 7 to avoid test error.In the decision tree figure the records are classified according to their point value. The labels X1, X2.......X13 are represented as attributes. In this dataset we are making use of 13 attributes. Each leaf node represents the class labels. Since it is a two classification problem it is classified as 1 for normal and 2 as sick. The decision tree [16, 17] classification analysis is shown in the Fig. 5. The last level of the decision tree shows the classified patterns.

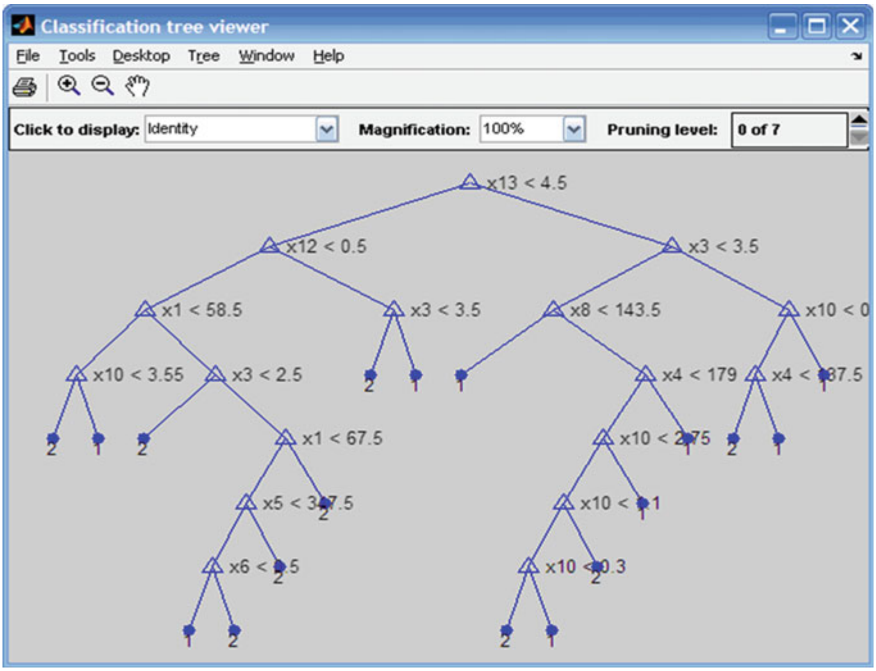


Fig. 5 Screen shots of multivariate decision tree

4 Result and Discussion

In this paper we are handling two class classification problem. One class is set of healthy people and another class is set of sick people. The attribute which classify the patterns are class label. Since we are handling with medical data, accuracy is a only consideration. Recent literatures [5, 15] shows the experimenting with the statically similar data set. If there is difference in data then the classification is very easy and we can get the accuracy in the higher side.

In our proposed evaluation system, after loading the CAD data, the data will be normalized by dividing each value of the attribute with the maximum value of that particular attribute (or column). This will lead to values between 1 and 2 which will be suitable for almost all the clustering and classification algorithm.

4.1 Evaluation Metrics Classifiers

The result of the classification is measured in terms of metrics [16, 20, 21] such as sensitivity, specificity and accuracy. The results are tabulated in the Tables 1, 2 and 3. These tables showing the comparative result [10] of classification analysis.

4.1.1 Sensitivity

Sensitivity measures the proportion of actual positives which are correctly identified as such (e.g. the percentage of sick people who are correctly identified as having the condition).

$$\text{Sensitivity} = \frac{\text{Number of True Positives}}{\text{Number of True Positives + Number of False Negatives}}$$

4.1.2 Specificity

Specificity measures the proportion of negatives, which are correctly identified (e.g. the percentage of healthy people who are correctly identified as not having the condition).

Table 1 Performance of K means algorithm

Training samples	K-means (sensitivity)	K-means (specificity)	K-means (accuracy)
1	74.45	75.00	74.45
2	73.72	76.25	73.72
3	54.01	85.63	54.01
Avg	67.39	78.96	67.39

Table 2 Performance of SOM

Training samples	SOM (sensitivity)	SOM (specificity)	SOM (accuracy)
1	72.26	83.13	72.26
2	59.85	58.13	59.85
3	59.12	56.88	59.12
Avg	63.74	47.09	63.74

Table 3 Performance of decision tree classifier with IDTC

Training samples	Decision tree classifier with IDTC (sensitivity)	Decision tree classifier with IDTC (specificity)	Decision tree classifier with IDTC (accuracy)
1	60.00	79.63	79.78
2	70.00	79.59	70.79
3	77.27	87.76	79.78
Avg	69.09	82.3	76.7

$$\text{Specificity} = \frac{\text{Number of True Negatives}}{\text{Number of True Negatives} + \text{Number of False Positives}}$$

4.1.3 Accuracy

Accuracy of a measurement system is the degree of closeness of measurements of a quantity to its actual (true) value.

$$\text{Accuracy} = \frac{\text{No of (True Positive + True Negatives)}}{\text{Number of True positives false Positives False Negative True Negative}}$$

In the Tables 1, 2 and 3 experiment results of classifiers are compared. The accuracy is important collective method which is directly showing the overall classification performance of the algorithm. According to the Accuracy [19] the decision tree based classifier is performed well. Figure 6 represents the performance of classifiers based on their accuracy.

In the Fig. 6 the x axis represents the classifiers and y axis represents the percentage of classification efficiency of the [11, 17, 22] classifiers.SOM is the poor performing algorithm among the all. After several, repeated analysis we made on the clustering and classification algorithms for classifying the CAD data, we came to the following conclusion. In the case of unsupervised machine learning algorithms SOM as well as the classical k-means, the performance of classification was very much depend up on the initial guess which is generally made during configuring these methods. For example, in the case of k-means the result depend on the initial guessed centroid. Hence leading to somewhat random [7, 9] results. Even though the results were random, the accuracy was not considerably good in almost all the cases. The decision tree with IDTC algorithm lead to better results.

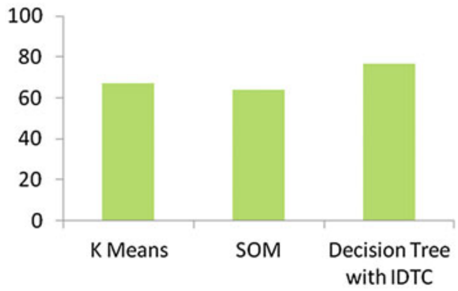


Fig. 6 Performance in terms of accuracy with the classifiers

5 Conclusion

Medical diagnosis is considered as a important task that needs to be carried out efficiently. The automation of this diagnosis would be highly beneficial. In this paper the problem of summarizing different algorithms of data mining used in the field of medical prediction are discussed. Three distinguished algorithms are compared with their performance to classify the multivariate data. The results obviously show the complex nature of data set restricts these algorithms from achieving better accuracy. If we, test the same algorithms with a normal synthetic data set with normal distribution, they provided good performance and better accuracy. In our review, out of three distinguished classifiers, the Decision tree with IDTC is outperformed. Our Future work will focus on improving the classification accuracy by implementing suitable methods.

References

1. Palaniappan RA (2008) Intelligent heartdisease prediction system using data mining techniques. IJCSNS Int J Comput Sci Netw Sec 8(8) (Aug)
2. Shantakumar BP, Kumaraswamy YS (2009) "Intelligent and effective heart attack prediction system using data mining and artificial neural network". Eur J Sci Res 31(4):642–656
3. Chen H et al (2010) Multi objective neural network ensembles based on regularized negative correlation learning. IEEE Trans Knowl Data Eng 22(12):1738–1743
4. Turker I, Serkin K et al. (2010) Evaluation of global and local training techniques over feed-forward neural network architecture spaces for computer-aided medical diagnosis, vol 37(12). Expert Systems with Applications Dec, Elsevier, Amsterdam
5. Setiawan NA, Venkatachalam PA et al (2009) Rule selection for coronary artery disease diagnosis based on rough set. Int J Recent Trends Eng Academy Publisher 2(5):1797–9617
6. Ceusters W (2000) Medical natural language understanding as a supporting technology for data mining in healthcare. In: Cios KJ (ed) Medical data mining and knowledge discovery Springer, Heidelberg, pp 32–60
7. Nalinipriya G, Kannan A, AnandhaKumar P (2012) A knowledgeable decision tree classifica tion model for multivariate heart disease data–A boon to healthcare ISICA, CCIS 316. Springer, Berling, pp 459–467
8. Hichem S (2011) Context-dependent kernels for object classification. IEEE Trans pattern anal mach intell Apr 33(4):699–708

9. Jian-Bo Y (2011) Determination of global minima of some common validation functions in support vector machine. IEEE Trans Neural Netw Apr 22(4):656–678
10. Colombet I, Ruelland A, Chatellier G, Gueyffier F (2000) Models to predict cardiovascular risk: comparison of CART, multilayer perceptron and logistic regression. In: Proceedings of AMIA Symp, pp 156–160
11. Detrano R “The Cleveland data”, VA Medical Centre, Long Beach and Cleveland Clinic Foundation
12. Windeatt T (2011) Embedded feature ranking for ensemble mlp classifiers. IEEE Trans Neural Netw 22(6):968–907 (Jun)
13. Boonjing V (2010) Comparing performance of logistic regression, decision trees and neural networks for classifying heart disease patients. In: proceedings of international conference on computer information system and industrial management Applications, pp 193–198
14. Pedro A (2011) Logistic regression by means of evolutionary radial basis function neural networks. IEEE Trans Neural Netw 22(2):246–278 (Feb)
15. Resul D, Ibrahim T et al. (2009) Effective diagnosis of heart disease through neural networks ensembles, vol 36(4) (May). Expert Systems with Applications, Elsevier, Amsterdam
16. Yanwei X, Wang J, Zhao Z, Gao Y (2007) Combination data mining models with new medical data to predict outcome of coronary heart disease. In: Proceedings international conference on convergence information technology, pp 868–872
17. Frank A, Asuncion A (2010) UCI machine learning respository. (online)
18. Avci E, Turkoglu I (2009) An intelligent diagnosis system based on principle component analysis and ANFIS for the heart valve diseases. J Expert Syst Appl 2(1):2873–2878
19. Bhatia S, Prakash P et al (2008) SVM based decision support system for heart disease classification with integer-coded genetic algorithm to select critical features. In: Proceedings of the World Congress on Engineering and Computer Science, WCECS ISBN: 978-988-98671-0-2
20. Deng Z (2010) Robust relief-feature weighting, margin maximization, and fuzzy optimization. IEEE Trans fuzzy syst 18(4) (Aug)
21. Alvarez G, Coiera E (2005) Interruptive communication patterns in the intensive care unit ward round. J Med Informatics 74(10):779–781
22. Cheng H et al (2010) Efficient algorithm for localized support vector machine. IEEE Trans Knowl Data Eng 22(4):381–389 (Apr)
23. Wang B (2011) ELITE: Ensemble of optimal input-pruned neural networks using trust-tech. IEEE Trans Neural Netw 22(4):96–107 (Jan)

Real Time Challenges to Handle the Telephonic Speech Recognition System

Joyanta Basu, Milton Samirakshma Bepari, Rajib Roy and Soma Khan

Abstract Present paper describes the real time challenges to design the telephonic Automatic Speech Recognition (ASR) System. Telephonic speech data are collected automatically from all geographical regions of West Bengal to cover major dialectal variations of Bangla spoken language. All incoming calls are handled by Asterisk Server i.e. Computer telephony interface (CTI). The system asks some queries and users' spoken responses are stored and transcribed manually for ASR system training. At the time of application of telephonic ASR, users' voice queries are passed through the Signal Analysis and Decision (SAD) Module and after getting its decision speech signal may enter into the back-end Automatic Speech Recognition (ASR) Engine and relevant information is automatically delivered to the user. In real time scenario, the telephonic speech contains channel drop, silence or no speech event, truncated speech signal, noisy signal etc. along with the desired speech event. This paper deals with some techniques which will handle such unwanted signals in case of telephonic speech to certain extent and able to provide almost desired speech signal for the ASR system. Real time telephonic ASR system performance is increased by 8.91 % after implementing SAD module.

Keywords Asterisk server · Interactive voice response · Transcription tool · Temporal and spectral features · Knowledge base

J. Basu (✉) · M. S. Bepari · R. Roy · S. Khan
Centre for Development of Advanced Computing, Kolkata, India
e-mail: joyanta.basu@cdac.in

M. S. Bepari
e-mail: milton.bepari@cdac.in

R. Roy
e-mail: rajib.roy@cdac.in

S. Khan
e-mail: soma.khan@cdac.in

Mohan S. and S. S. Kumar (eds.), *Proceedings of the Fourth International Conference on Signal and Image Processing 2012 (ICSIP 2012)*, Lecture Notes in Electrical Engineering 222,
DOI: 10.1007/978-81-322-1000-9_38,

1 Introduction

Modern human life is totally dependent on technology and along with these devices become more and more portable like mobile, PDAs, GPRS etc. Beside this, there is also a growing demand for some hands free Voice controlled public purpose emergency information retrieval services like Weather forecasting, Road-Traffic reporting, Travel enquiry, Health informatics etc. accessible via hand-held devices (mobiles or telephones) to fulfill urgent and on the spot requirements. But real life deployment of all these applications involves development of required modules for voice-query based easy user interface and quick information retrieval using mobiles. In fact, throughout the world the number of telephone users is much higher than that of the PCs. Again human voice or speech is the fastest communication form in our daily busy schedule that further extends the usability of such voice enabled mobile applications in emergency situations. In such a scenario, speech-centric user interface on smart hand-held devices is currently foreseen to be a desirable interaction paradigm where Automatic Speech Recognition (ASR) is the only available enabling technology.

Interactive Voice Response (IVR) systems provide a simple yet efficient way for retrieving information from computers in speech form through telephones but in most of the cases users still have to navigate into the system via Dual Tone Multiple Frequency (DTMF) input and type their query by telephone keypad. A comparative study by Lee and Lai [1] revealed that in spite of occasionally low accuracy rates, a majority of users preferred interacting with the system by speech modality as it is more satisfying, more entertaining, and more natural than the touch-tone modality which involves the use of hands, quite time consuming and require at least the knowledge of English alphabets. Furthermore, a variety of ASR system architectures [2] have been implemented ranging from server based implementations accessed by the device over wireless network to recognizers embedded in the local processor associated with specific device [3].

Present paper addresses some real time challenges to handle telephonic ASR application to provide better result. And also provides a clear picture of the above tasks in a well planned and sequential manner aiming towards the development of an IVR application in spoken Bangla language. The methodology is language independent and can easily be adapted to any applications of similar type.

2 Motivation of the work

A practical IVR system should be designed in such a way that it should be capable of handling real time telephony hazards like channel drop, clipping, speech truncation etc. It should also provide robust performance considering following issues:

1. Speaker-Variability: Handle speech from any arbitrary speaker of any age i.e., it would be a speaker-independent ASR system.
2. Pronunciation/Accent Variability: Different pronunciations, dialectical variations and accents within a particular Indian language.
3. Channel Variability: Different channels such as landline versus cellular and different cellular technologies such as GSM and CDMA.
4. Handset Variability: Variability in mobile handsets due to differences in spectral characteristics.
5. Different Background Noise: Various kinds of environmental noise, so that it is robust to real-world application.

Considering the above said requirements, telephonic ASR is being designed in such a way that, to-some-extent it can meet the above mentioned capabilities. Speech data are mainly collected from all the geographical regions where native Bangla language speaking population is considerably high. The collected speech data is then verified and used for ASR training. The reason behind choosing a large geographical area for data collection is to cope up with the problem of speaker variability, accentual variability. Additionally, various issues regarding the telephonic channel such as channel drop or packet lost during transmission, handset variability, service provider variability, various types of background noise such as cross-talk, vehicle noise etc. have been observed, analyzed and estimated efficiently from the collected speech data and modeling of those can improve ASR performance. These issues will not only help us to improve the system performance effectively, but also provide us very good research motivation on other telephonic applications.

3 Brief Overall System Overview

Telephonic ASR system is designed such a way, that users get the relevant information in a convenient manner. First the system will give the user a language preference (within Hindi, Bangla and Indian English) and then onwards each time a directed question is asked, and the user would reply it with appropriate response from a small set of words. System is composed of three major parallel components. They are IVR server (hardware and API), Signal Processing Blocks and Information Source. Figure 1 represents an overall block diagram of the system.

3.1 IVR Hardware and API

As shown in Fig. 2 the Interactive Voice Response (IVR) consists of IVR hardware (generally a Telephony Hardware), a computer and application software running on the computer. The IVR hardware is connected parallel to the telephone line.

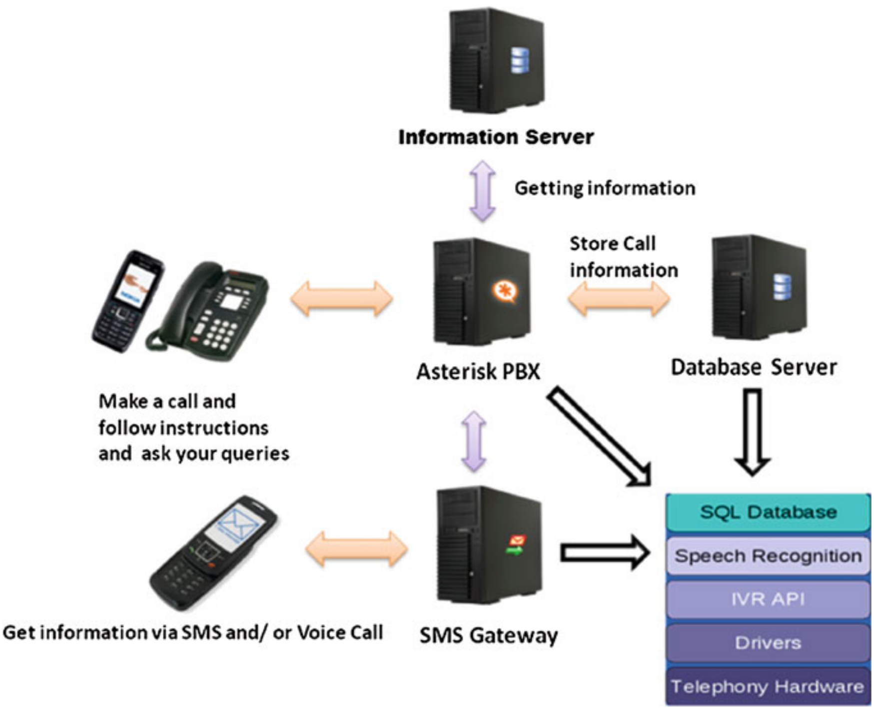


Fig. 1 Block diagram of telephonic ASR system

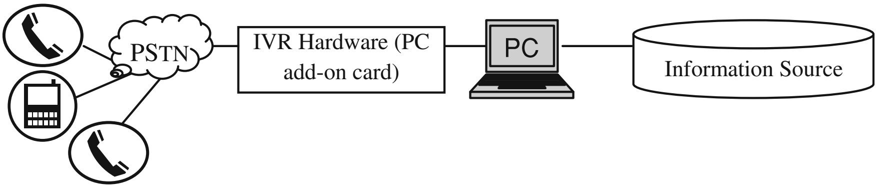


Fig. 2 Interactive voice response system

The functionality of the IVR hardware is to lift the telephone automatically when the user calls, recognize the input information (like dialed digit or speech) by the user, interact with computer to obtain the necessary information, then convert the information into speech form and also convert the incoming speech into digital form and store it in the computer.

In development of telephonic ASR system, Asterisk [4, 5] is used here as an open source IVR Server, converged telephony platform, which is designed primarily to run on Linux. It support VoIP protocols like SIP, H.323; interfaces with PSTN Channels, supports various PCI Cards, and also open source Drivers and Libraries are available.

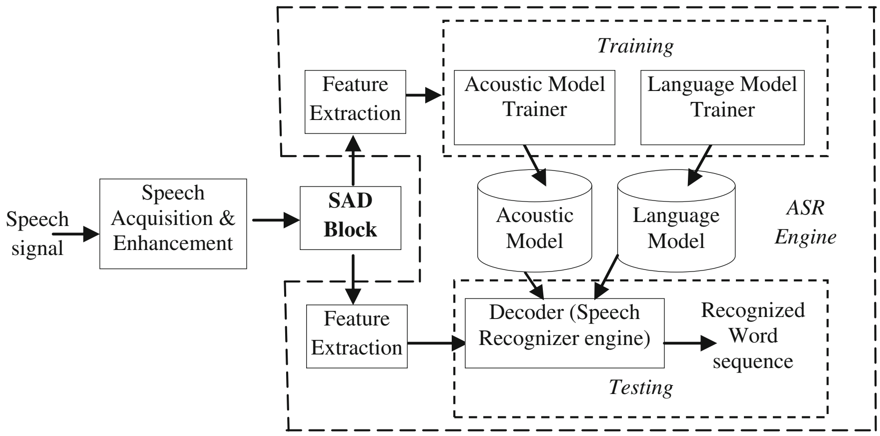


Fig. 3 Basic block diagram of signal processing blocks including automatic speech recognition system

3.2 Signal Processing Block

This block consists of three major blocks namely Speech Acquisition and Enhancement module, Signal Analysis and Decision module and ASR Engine. Block diagram of such Signal Processing Block is shown in Fig. 3.

3.2.1 Speech Acquisition and Enhancement Module

The first block, which consists of the acoustic environment plus the transduction equipment, can have a strong effect on the generated speech representations because additive noise, room reverberation, recording device type etc. are associated with the process. A speech enhancement module suppresses the above effects so that the incoming speech can easily be recognized in heavy perturbed conditions.

3.2.2 Signal Analysis and Decision Module

In this module all incoming speech waveform analyzed for the valid speech signal or not. This is the very important module of this system. This module extracts different temporal features like Zero Crossing Rate (ZCR), Short Time Energy (STE) and Spectral features like formant analysis etc. And also using some predefined Knowledge Base (KB) this module gives some decision valid information regarding incoming speech signal. After this module system takes some decision whether users need to re-record speech signal or not. If re-recording is not required then recorded speech signal go through ASR engine for decoding the signal. Detail of this module is given in the Sect. 4. In this paper we are mainly looking towards the performance of this module.

3.2.3 ASR Engine

The basic task of Automatic Speech Recognition (ASR) is to derive a sequence of words from a stream of acoustic information. Automatic recognition of telephonic voice queries requires a robust back-end ASR system. CMU SPHINX [6], an open Source Speech Recognition Engine is used here which typically consists of Speech Feature Extraction module, Acoustic Model, Language Model, Decoder [7].

3.3 Information Source

Repository of all relevant information is known as a trusted Information Source, and design architecture of the same typically depends on type of information. In current work dynamic information is mainly refereed from trusted information source or online server and all other information which does not change very much during a considerable period of time are kept in local database using web crawler. As an example, PNR record, seat availability and running status of any train are typically dynamic information for Travel Domain. System response of any query on dynamic information must ensure delivery of latest information. To accomplish this objective a specific reference made to trusted information source. On other hand the information like train schedule, train name, train number are fetched and stored periodically in the local database. This approach ensures quick delivery of information.

4 Signal Analysis and Decision Module

This is one of the challenging modules of this entire telephonic ASR system. Where all speech signals is analyzed with temporal and spectral features and with the help of previous knowledge it will decide what to do next with stored speech signal. Here in this module currently four types of erroneous signals are considered. Those are occurs due to below problems:

1. Telephonic channel related problem
2. Signals sometimes truncated from the beginning and/or end of the signal,
3. Sometimes silence signals are coming
4. Lots of background noise like air flow, sudden noise etc. and cross talks are there.

This module consists of Feature extraction module, Decision module and Knowledge base (KB). Figure 4 shows the basic speech analysis and decision module blocks.

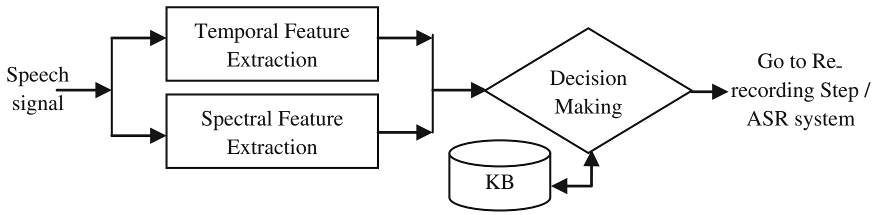


Fig. 4 Speech analysis and decision module blocks

4.1 Temporal Feature Extraction

In this work, various temporal features such as Zero Crossing Rate (ZCR) and Short-Time Energy (STE) are used.

4.1.1 Zero Crossing Rate

It is the rate of sign changes along the signal, i.e. the rate at which signal changes from positive to negative or vice versa. ZCR is more for unvoiced speech than the Voiced speech. It is defined as,

$$zcr = \frac{1}{N-1}\sum_{n=1}^{N-1} \mathrm{II}\{s(n)s(n-1) < 0\},$$

where s(n) is speech signal of length N and indicator function $\mathrm{II}\{A\}$ is 1 if its argument A is true and 0 otherwise.

4.1.2 Short-Time Energy

In this work, STE is used to determine the energy of voiced and unvoiced regions of the signal. STE can also be used to detect the transition from unvoiced to voiced regions and vice versa. The energy of voiced region is greater than unvoiced region. The equation for STE is given by,

$$E_n = \sum_{m=-\infty}^{\infty} \{s(m)\}^2 h(n-m),$$

where s(m) is the signal and h(n-m) is the window function.

4.2 Spectral Feature Extraction

In this work we are only extracting the formant parameters from the speech signal. Formants of voiced signal is clearly visible than unvoiced signal. The center frequency of the lowest resonance of the vocal tract, called first formant frequency or F1, corresponds closely to the articulatory and/or perceptual dimension of vowel height (high vs. low vowels, or close vs. open vowels). The vowel classification has been carried out by measuring formant frequencies (F1, F2, F3).

4.3 Knowledge Base

KB module gathered information from the transcribed speech data which is required for ASR system training. Transcription tool [8] has been designed for offline transcription of recorded speech data, such that all transcriptions during data collection can be checked, corrected and verified manually by human experts. Automatic conversion of text to phoneme (phonetic transcription) is necessary to create pronunciation lexicon which will help the ASR System training.

The methodology for Grapheme to Phoneme (G2P) conversion in Bangla is based on orthographic rules. In Bangla G2P conversion sometimes depends not only on orthographic information but also on Parts of Speech (POS) information and semantics [9]. G2P conversion is an important task for data transcription. From where, many information were gathered regarding telephonic speech data. At the time of transcription we have to give some transcription remark tags and

Table 1 Description and measurement of remarks

Wave remarks (Wr)	Description
A_UTTR (Amplitude)	Amplitude (Partly or fully) will be modified
C_UTTR (Clean)	Speech utterance may contain some non overlapping non-speech event
Transcription remarks (Tr)	**Description**
CLPD_UTTR (Clipped)	Clipping of speech utterance
CPC_UTTR (Channel problem consider)	Channel drop occurs randomly in silence region which have not affect speech region
CPR_UTTR (Channel problem reject)	Some word or phonemes dropped randomly
I_UTTR (Improper)	In this case the utterance is slightly different than prompt in phoneme level
MN_UTTR	Noise within speech
MP_UTTR	Reasonable silence (pause) within speech
R_UTTR (Reject)	In this case speech signal is wrongly spelt or may be too many noise or may be non-sense words, can't be able to understand
S_UTTR (Silence)	No speech utterance present
TA_UTTR (Truncate accept)	Truncation of not so significant amount (may be one or two phoneme) speech utterance
TR_UTTR (Truncate reject)	Truncation of significant amount speech utterance
W_UTTR (Wrong)	In this case the utterance is totally different from the corresponding prompt

also noise tags. It's totally human driven task. Descriptions and measurement of the remarks are given in Tables 1 and 2 shows the different types of noise tags.

From Table 1 it has been seen that S_UTTR, CPR_UTTR, TR_UTTR and R_UTTR are the rejection tag set. Other tag set may be accepted and considered at the time of ASR training. Now all four rejection tags consider as knowledge from offline data transcription. And main objective of SAD module is to extract all those rejection remarks from the incoming speech signal on-the-fly using this KB.

5 Observations and Results

For this work we have collected almost 60 h telephonic speech data from nineteen districts of West Bengal and transcribed manually. This transcribed data helps to build up the KB. Table 3 represents the distribution of totally collected speech data according to the variations of Speakers' Gender, Age, Education Qualification, Recording Handset model, Service provider and Environment in terms of

Table 2 Types of noise tags

Tag	Explanation / examples
<air>	Air flow
<animal>	Animal sound
<bang>	Sudden (impulsive) noise due to banging of door
<beep>	Telephonic beep sound
<bird>	Sound of bird
<bn>	General background noise
 	Breath noise
<bs>	Background speech (babble)
<bsong>	Background song
<bins>	Background instrument
<burp>	Burp
<cough>	Cough
<cry>	Children cry
<ct>	Clearing of throat
<horn>	Horn noise of vehicles
<ht>	Hesitation
<laugh>	Laugher
<ln>	Line noise
<ls>	Lip smack
<ns>	Hiccups, yawns, grunts
<pau>	Pause or silence
<ring>	Phone ringing
<sneeze>	Sneeze
<sniff>	Sniff
<tc>	Tongue click
<vn>	Vehicle noise

Table 3 Variations in collected speech data

Gender (in %)	Male	66
	Female	34
Age (in %)	Child: 0–15	11
	Adult: 15–30	56
	Medium: 30–50	28
	Senior: 50–99	5
Qualification (in %)	Primary	12
	Secondary	46
	Post-secondary	38
	Others	4
Handset model (in %)	Nokia	46
	Samsung	20
	LG	5
	Reliance	2
	Sony	2
	Others	25
Service provider (in %)	BSNL	11
	Airtel	10
	Vodafone	29
	Reliance	3
	Aircel	11
	MTS	15
	IDEA	17
	Others	4
Environment (in %)	Noise	2
	Clean	85
	Babble	9
	Music	4

percentage of occurrences in each criteria. Figure 5 presents the true picture of the nature of real world speech data after being checked and remarked by human experts with the help of the Transcription tool. Table 4 described the percentage of occurrence of major noise tags.

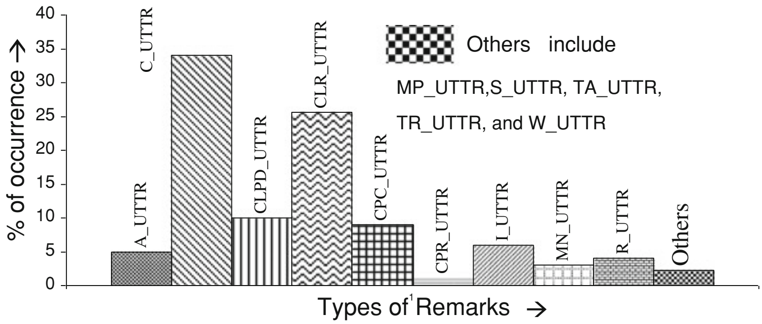


Fig. 5 Percentage of occurrence of different types of remarks

Table 4 Percentage of occurrence of the major noise tag set

Tag	% of occurrence
<air>	7.92
<bang>	1.09
<beep>	5.59
<bird>	5.22
<bn>	17.91
 	0.4
<bs>	15.9
<cough>	6.1
<ct>	0.03
<horn>	1.02
<laugh>	8.08
<ln>	6.06
<ring>	6.08

Figure 6 shows the observation result of CPR_UTTR, S_UTTR and TR_UTTR speech utterances, STE and ZCR plot. It has been observed from the figures that (1) In case of Channel Problem (i.e. CPR): signals ends with high STE and high SCR value may be at the end of the signal or sometimes occurs at the beginning of the signal. It happens because users some time delayed speaking within specific

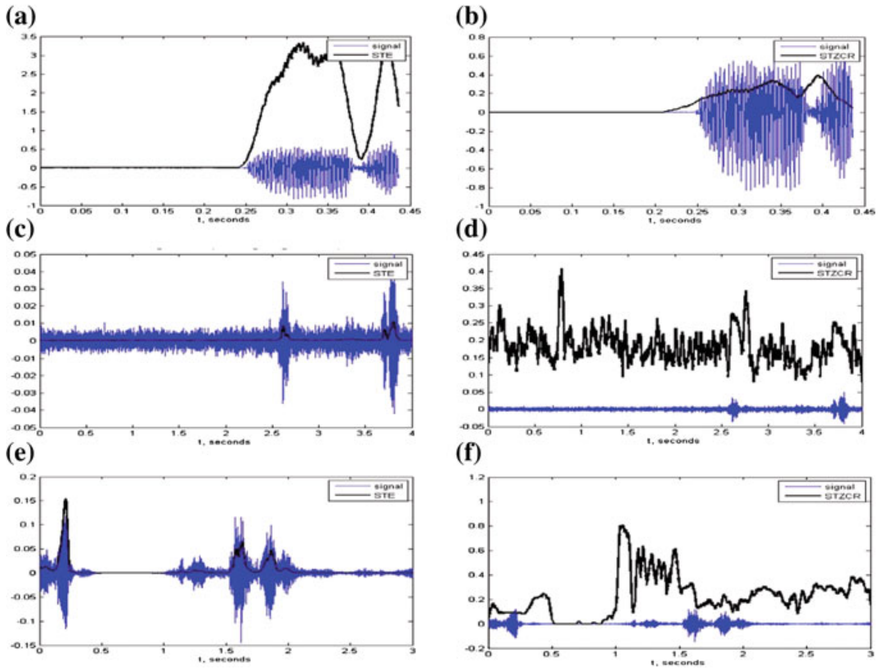


Fig. 6 CPR_UTTR. **a** Signal vs. STE. **b** Signal vs. ZCR; S_UTTR. **c** Signal vs. STE. **d** Signal vs. ZCR; TR_UTTR. **e** Signal vs. STE. **f** Signal vs. ZCR

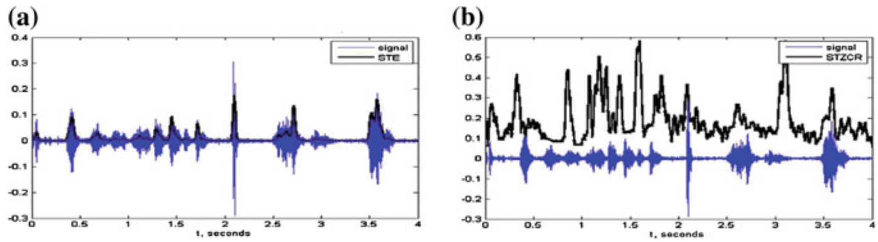


Fig. 7 R_UTTR. **a** Signal vs. STE. **b** Signal vs. ZCR

Table 5 Output of signal analysis module and its decision

Number of incoming calls	Number of utterances	CPR_UTTR	S_UTTR	TR_UTTR	R_UTTR
887	10,327	659	795	248	458

time span or sometimes starts so early. (2) In case of S_UTTR: signals are generally channel noise and mostly unvoiced sounds, that's why high ZCR and low STE observed almost entire time span of the recording. (3) In case of TR_UTTR: sometimes channel packets have been lost due to may be bad network strength of the users or may be type of handset. It's very common problem for the telephonic applications. It has been observed that ZCR value change suddenly from low to high or high to low and same thing for STE also. Formant analysis also has been done for all these rejection cases. To find out S_UTTR automatically from the signal formant may be the one good spectral feature. But for other cases of rejection, formant analysis is not so well.

Figure 7 shows the one of the observation of R_UTTR, where it has been seen that many voiced and unvoiced regions are there, but those are basically background speech, not the expected reply from the uses. And also we have seen that natural STE and ZCR plots are there. So, it's really difficult to automatic extract R_UTTR always. But sometimes, when overall STE is below the expectation level of the incoming speech signal then Speech Analysis module marks it as a R_UTTR. Basically a performance of automatic extraction of R_UTTR depends on the background noise or unwanted speech. More work is going on this rejection type.

From the above observations, signal analysis module are designed for telephonic ASR system and tested in real life scenarios. On-the-fly pattern of rejection remarks extraction is the main objective of signal analysis module. We have analyzed 887 numbers of incoming calls of users and total utterances are 10,327 numbers. Table 5 shows result of automatic extraction of the rejection remarks except R_UTTR. To find out R_UTTR remarks system desires some manual intervention.

Table 6 shows the performance accuracy of the signal analysis and decision blocks. It has been observed that S_UTTR, TR_UTTR percentage accuracy is good than CPR_UTTR. But automatic extraction of R_UTTR is not satisfactory.

Table 7 shows the Telephonic ASR performance on using Signal Analysis and decision module and without using Signal Analysis and decision module. It has

Table 6 Accuracy of signal analysis and decision module

Types of rejection remarks	% of correct decision
CPR_UTTR	65.85
S_UTTR	86.98
TR_UTTR	87.26
R_UTTR	32.65

Table 7 Telephonic ASR system accuracy

Number of utterances	Without using signal analysis and decision module	With using signal analysis and decision module
10327	62.68%	71.59%

been seen that with the use of Signal analysis and decision module percentage of accuracy improved by 8.91% which is really encouraging for next level of research on this work.

6 Conclusion

In the proposed work, a detail design of signal analysis and decision module has been described to cope up the real time challenges of telephonic automatic speech recognition system. Here in this work try to address various issues regarding the telephonic channel such as channel drop or packets lost during transmission and try to handle some issues. We have seen that automatic extraction of CPR_UTTR, S_UTTR and TR_UTTR by SAD module is encouraging. But automatic extraction of R_UTTR is not so easy. Currently more research is going on for this particular utterance type. It has been also observed that with the help SAD module telephonic ASR accuracy percentage increased by 8.91, which is really encouraging for the researchers. More importantly, this kind of real voice based information retrieval application useful especially to the people having no access to computers and Internet, people who may not have the required computer skills or even reading/writing abilities and also the visually challenged personal. After successful completion of the present work, it will enable development of similar speech-based access systems for other (like Medical, tourism, transport, Emergency services) public domain applications.

References

1. Lee K-M, Lai J (2005) Speech vs. touch: a comparative study of the use of speech and DTMF keypad for navigation. International Journal of Human Computer Interaction IJHCI, vol 19(3)
2. Furui S (2000) Speech recognition technology in the ubiquitous/wearable computing environment. In: Proceedings of the international conference on acoustics speech and signal processing, pp 3735–3738

3. Maes SH, Chazan D, Cohen G, Hoory R (2000) Conversational networking: conversational protocols for transport, coding, and control. In: Proceedings of the international conference on spoken language processing
4. Gomillion D, Dempster B Building telephony system with asterisk. ISBN: 1-904811-15-9, Packet Publishing Ltd
5. Meggelen JV, Madsen L, Smith J Asterisk: the future of telephony, ISBN-10: 0-596-51048-9, ISBN-13: 987-0-596-51048-0, O'REILL
6. http://www.speech.cs.cmu.edu/
7. Basu J, Khan S, Roy R, Bepari MS (2011) Designing voice enabled railway travel enquiry system: an IVR based approach on bangla ASR. ICON 2011, Anna University, Chennai, India, pp 138–145
8. Basu J, Bepari MS, Roy R, Khan S (2012) Design of telephonic speech data collection and transcription methodology for speech recognition systems. FRSM 2012, KIIT, Gurgaon, pp 147–153
9. Basu J, Basu T, Mitra M, Das Mandal S (2009) Grapheme to Phoneme (G2P) conversion for bangla. O-COCOSDA international conference, pp 66–71

Ranking Sector Oriented Sense with Geographic Protocol

Rajalakshmi Dheenadayalan and Sanoj Subramanian

Abstract There are several multicast routing protocols in MANET have been shown to large overhead due to high networks dynamic topology. In MANET, group communications are different form. Multicast is a fundamental service for supporting information exchanges and collaborative task execution among a group of users. To overcome these limitations the protocol termed as Stateless and Secure On Demand Multicast Routing Protocol (SODMRP) proposes a novel Zone based Geographic multicast protocol. Several virtual architectures are used in this protocol that avoids the difficulty in group membership management, packet forwarding and maintenance. Here, the GPRS are used to provide the location information of group members that will reduce the joining delay. The position information is efficiently used to reduce the route searching and tree structure maintenance. The scalability and the efficiency of proposed protocol are evaluated through simulations and quantitative analysis…Compared with existing protocol like SPBM our novel results demonstrate that proposed protocol has high packet delivery ratio, low control overhead and multicast group joining delay under all test scenarios with different moving speeds, node densities, number of groups and network sizes.

Keywords Routing · Wireless networks · Mobile ad hoc networks · Location · Geographic multicast · Multicast · Protocol

R. Dheenadayalan (✉) · S. Subramanian
Assistant Professor, Department of Information Technology,
Shri Angalamman College of Engineering and Technology, Siruganur,
Trichy 621 105, Tamilnadu, India
e-mail: rajisacet@gmail.com

S. Subramanian
e-mail: sanojsubramanian@gmail.com

Mohan S. and S. S. Kumar (eds.), *Proceedings of the Fourth International Conference on Signal and Image Processing 2012 (ICSIP 2012)*, Lecture Notes in Electrical Engineering 222,
DOI: 10.1007/978-81-322-1000-9_39,

1 Introduction

The technology and popularity of the Internet have grown, applications that require multicasting (e.g., video conferencing) are becoming more widespread. Another interesting recent development has been the emergence of dynamically reconfigurable wireless ad hoc networks to interconnect mobile users for applications ranging from disaster recovery to distributed collaborative computing [1]. Multicast plays a key role in ad hoc networks because of the notion of teams and the need to show data/images to hold conferences among them Multicast tree structures are fragile and must be readjusted continuously as connectivity changes. Furthermore, typical multicast trees usually require a global routing substructure such as link state or distance vector. The frequent exchange of routing vectors or link state tables, are triggered by continuous topology changes, yields excessive channel and processing overhead. Limited bandwidth, constrained power, and mobility of network hosts make the multicast protocol design particularly challenging [16].

Although it is important to support multicast in a mobile ad hoc network (MANET) [13], which is often required by military and emergency applications, there is a big challenge to design a reliable and scalable multicast routing protocol in the presence of frequent topology changes and channel dynamics [7]. Many efforts have been made to develop multicast protocols for MANETs. These include conventional tree-based protocols and mesh-based protocols [18]. The tree-based protocols However, it is very difficult to maintain the tree structure in mobile ad hoc networks, and the tree connection is easy to break and the transmission is not reliable. The mesh-based protocols (e.g., FGMP [11], Core-Assisted Mesh protocol [18], ODMRP [10]) are proposed to enhance the robustness with the use of redundant paths between the source and the set of multicast group members incurs higher forwarding overhead. There is a big challenge to support reliable and scalable multicast in a MANET with these topology-based schemes, as it is difficult to manage group membership, find and maintain multicast paths with constant network topology changes [8].

Additionally, there is a need to efficiently manage the membership of a potentially large group, obtain the positions of the members, and transmit packets to member nodes that may be located in a large network domain and in the presence of node movements. The existing small-group based geographic multicast protocols normally address only part of these problems [3].

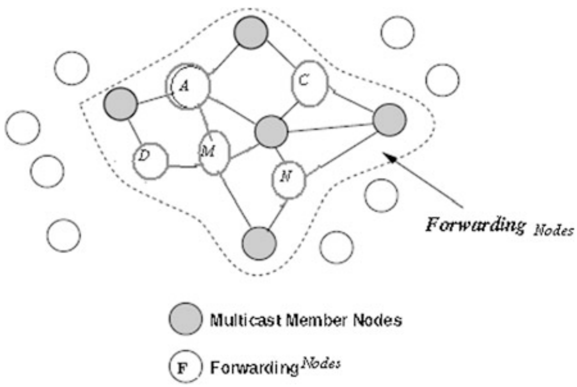
In this paper, the protocol is designed to be simple, thus it can operate more efficiently and reliably. We introduce several virtual architectures for more robust and scalable membership management and packet forwarding in the presence of high network dynamics due to unstable wireless channels and frequent node movements [2]. Both the data packets and control messages will be transmitted along efficient tree-like paths, however, different from other tree-based protocols, there is no need to explicitly create and maintain a tree structure. A robust virtual-tree structure can be formed during packet forwarding with the guidance of node positions. Furthermore, it makes use of position information to support reliable packet forwarding. The protocol is designed to be comprehensive and self-contained. Instead of addressing only a specific part of the problem, it introduces a zone-based scheme to efficiently handle the group membership management, and takes advantage of the membership management structure to efficiently track the locations of all the group members without resorting to any external location server [14]. The zone structure is also formed virtually and the zone where a node is located can be calculated based on the node position and a reference origin [6]. Different from conventional cluster structures, there is no need to involve a complicated scheme to create and maintain the zone. To avoid the need of network-wide periodic flooding of source information [9], we introduce Root to track the positions and addresses of all the sources in the network.

2 Literature Review

Effective location guided tree construction algorithms for the small group multicast scheme in mobile ad hoc networks (MANET) is based on packet encapsulation [17]. The constructed packet distribution tree resides on top of the underlying unicast routing protocol, and spans over all members of a multicast group. The algorithms include a hybrid location update mechanism to exchange location information among a group of nodes. This scheme has the drawback of Enormous redundant packet deliveries: congestions, collisions, etc. and the Performance goes down drastically when there is more than one source in the transmission.

The key characteristic of AM Route is its usage of virtual mesh links to establish the multicast tree. Therefore, as long as routes between tree members exist via mesh links, the tree need not be readjusted when network topology changes [18]. Nonmembers do not forward data packets and need not support any multicasting protocol. Thus, only the member nodes that form the tree incurs processing and storage overhead. AM Route relies on an underlying unicast protocol to maintain connectivity among member nodes and any unicast protocol can

Fig. 1 ODMRP group concept



be used. The major disadvantage of the protocol is that it suffers from temporary loops and creates non-optimal trees when mobility is present (Fig. 1).

The SPBM is a quad tree-based protocol and ODMRP is a mesh-based on-demand non-geographic multicast protocol, and takes a soft-state approach to maintain multicast group members [10]. A multicast source broadcasts Join-Query messages to the entire network periodically. An intermediate node stores the source ID and the sequence number, and updates its

Routing table with the node ID (i.e. backward learning) from which the message was received for the reverse path back to the source. A receiver creates and broadcasts a Join Reply to its neighbors, with the next hop node ID field filled by extracting information from its routing table. The neighboring node whose ID matches that in the message broadcasts its own Join Table built upon matched entries. This whole process constructs (or updates) the routes from sources to receivers and builds a mesh of nodes, the forwarding group.

The mesh structure in ODMRP has more redundancy when more nodes join the multicast group and will provide more robust delivery paths. In ODMRP, all the mobile nodes are involved in the periodic flooding of JOIN QUERY, which results in a higher normalized control overhead [10]. In SPBM, the proactive multi-level control message flooding causes much more unnecessary overhead when the group size is small relative to the total number of mobile nodes.

AMRIS establishes a shared tree for multicast data forwarding. Each node in the network is assigned a multicast session ID number [5]. The ranking order of ID numbers is used to direct the flow of multicast data. Like ODMRP, AMRIS does not require a separate unicast routing protocol. AMRIS detects link disconnection by a beaconing mechanism. If no beacons are heard for a predefined interval of time, the node considers the neighbor to have moved out of radio range. If the former neighbor is a parent, the node must rejoin the tree by sending a JOIN-REQ to a new potential parent. If the node fails to join the session or no qualified neighbors exist, it performs the Branch Reconstruction process data forwarding in done by the nodes in the tree. Only the packets from the registered parent or registered child are forwarded. Hence, if the tree link breaks, the packets are lost until the tree is reconfigured.

3 Proposed Methodology

Tree-based protocols construct a tree structures for more efficient forwarding of packets to all the group members. Mesh-based protocols expand a multicast tree with additional paths that can be used to forward multicast data packets when some of the links break. A topology-based multicast protocol generally has the following three inherent components that make them difficult to scale:

- Group membership management.
- Creation and maintenance of a multicast structure.
- Multicast Data forwarding.

As the focus of our paper is to improve the scalability of location-based multicast, a comparison with topology-based protocols is out of the scope of this work. However, we note that at the similar mobility and system set-up, it has a much higher packet delivery ratio.

- To reduces the control overhead and increases the reliability and scalability of the protocol, the Proposing stateless distribution schemes that have to send the data packets and control messages along efficient virtual tree paths without the need of explicitly building and maintaining a tree-structure as in conventional tree-based multicast protocols.
- To make use of the position information to design a Scalable and reactive zone-based scheme for efficient Membership management, which allows a node to join and leave a group quickly.
- Efficient location search of multicast group members, by combining the location service with the membership management to avoid the need and overhead of using a separate location server.
- Source Home to track the addresses and positions of the sources, to avoid network-wide periodic flooding of source information.

Designing schemes to handle the empty-zone problems for both member zones and the Source Home are critical in designing a zone-based protocol.

Many issues need to be addressed to make the protocol fully functional and robust. The issues related to zone management include: the strategy for electing a zone leader on-demand and maintaining the zone leader during mobility, the handling of empty zone problem, the scheme for Source Home construction and maintenance, and the need to reduce packet loss during node moving across zones [6]. The issues related to packet forwarding include: the scheme for virtual tree construction without the need of storing and tracking tree-state information, and the reliable transmissions of control and multicast data packets without resorting to an external location server (Fig. 2).

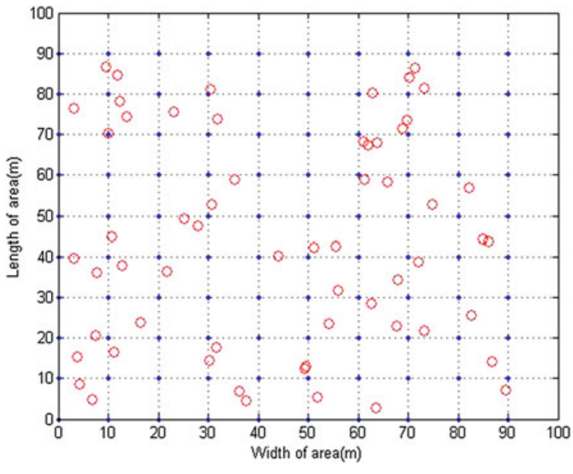


Fig. 2 Zone construction with node deployment

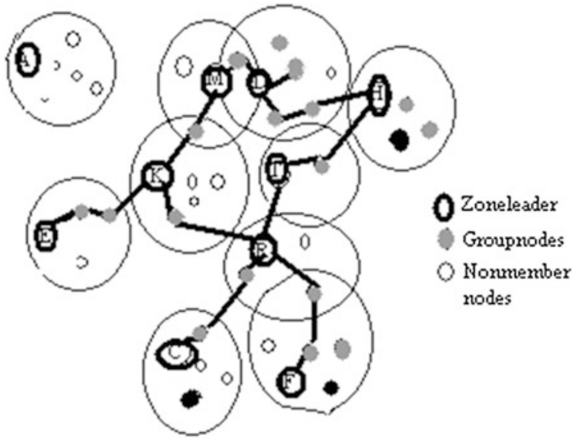


Fig. 3 Zone based communication

The forwarding of data packets and most control messages is based on a geographic unicast routing protocol [15]. In our performance study, we implemented GPSR as an underlying unicast protocol to support the packet transmissions. The protocol, however, does not depend on a specific geographic unicast routing protocol. Some of the notations to be used are:

Pos: A mobile node's position coordinates (x, y).

Zone: The network terrain is divided into square zones. We will also study the impact of zone size on the performance of the protocol.

mZone (non mZone): Member (Non member) zone, a zone with (without) group (Fig. 3).

It supports a two-tier membership management and forwarding structure. At the lower tier, a zone structure is built based on position information and a leader is elected on demand when a zone has group members [4]. A leader manages the group membership and collects the positions of the member nodes in its zone. At the upper tier, the leaders of the member zones report the zone membership to the sources directly along a virtual reverse-tree-based structure. It contains the following categories,

A. Source side information
B. Lowest level membership management
C. Network level Management
D. Creation of NEW SESSION from the source to group of nodes.
E. Multicast Packet Delivery.
F. Cost for zone building and geographic routing

4 Performance Evaluation

The zone is virtual and determined by each node based on its position and the reference origin, without the need of extra signaling messages. The leader information is distributed with a flag inserted in the beacon messages of the underlying geographic unicast routing protocol [15]. Therefore, the per node cost of the zone building and geographic routing is impacted by the beaconing frequency $1 = \text{Intvalmin}$ The tree construction process is associated with the multicast session initiation and termination, and the member joining and leaving the multicast tree.

Focus on the studies of the scalability and efficiency of the protocol under the dynamic environment and the following metrics were used for the multicast performance evaluation:

1. Packet delivery ratio: The ratio of the number of packets received and the number of packets expected to receive. Thus for multicast packet delivery, the ratio is equal to the total number of received packets over the number of originated packets times the group size.
2. Normalized control overhead: The total number of control message transmissions divided by the total number of received data packets. Each forwarding control message was counted as one transmission. Different from ODMRP and SPBM, the proposed protocol is based on some underlying geographic unicast routing protocol which involves use of periodic beacons.
3. Normalized data packet transmission overhead: The ratio of the total number of data packet transmissions and the number of received data packets.
4. Joining delay: The average time interval between a member joining a group and its first receiving of the data packet from that group. To obtain the joining delay, the simulations were rerun with the same settings except that all the members joined groups after the source began sending data packets.

First compare the performance of ODMRP, SPBM and SODMRP with the variation of moving speed and node density, we then study the scalability of the three protocols with the change of group size and network size, then evaluate the performance of the protocol with the following factors,

- Impact on mobility
- Impact on node density
- Impact on zone size
- Impact on group size

In summary, SODMRP can perform much better than SPBM and ODMRP in a large network, and has a significantly higher delivery ratio, lower control overhead, and lower joining delay due to its virtual and reliable membership management and transmission infrastructures.

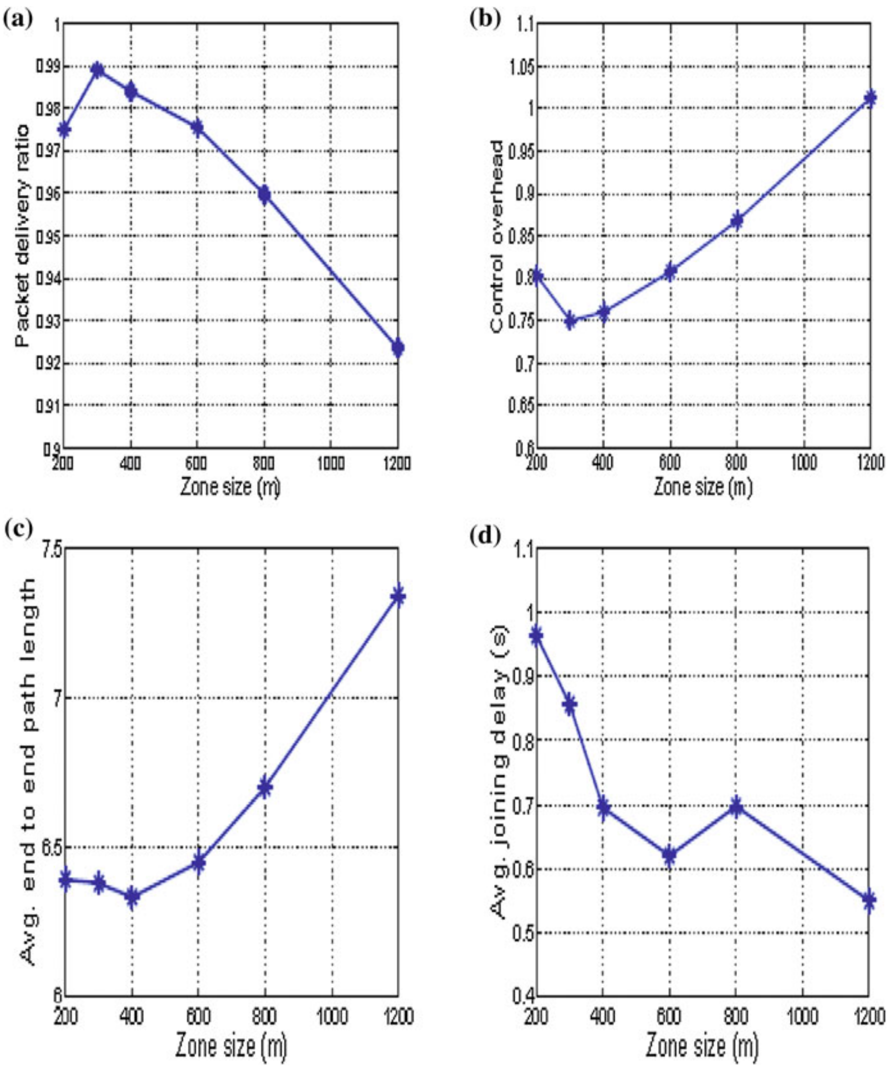


Fig. 4 Performance vs. zone size

5 Conclusion

In this paper, we have designed a zone based geographic multicast protocol SODMRP for MANET. In SODMRP, scalable and stateless virtual transmission structures are used for simple management and robust forwarding. Both data packets and control messages are transmitted along efficient tree-like paths without the need of explicitly creating and maintaining a tree structure. Scalable

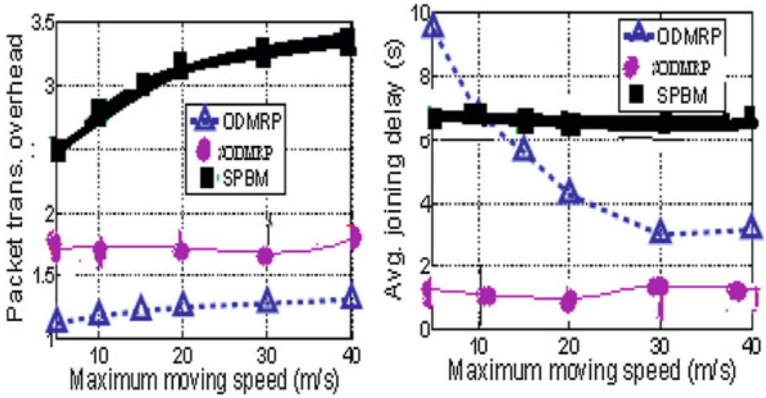
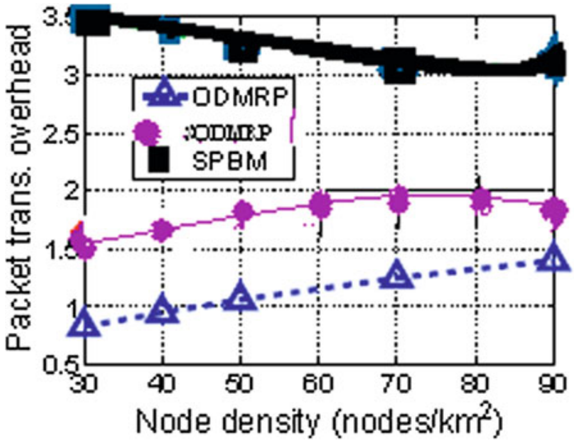


Fig. 5 Performance vs. maximum moving speed

Fig. 6 Performance vs. node density



membership management is achieved through a virtual-zone-based two-tier infrastructure. A Root node is defined as the source to track the locations and addresses of the multicast sources to avoid the periodic network-wide flooding of source information, and the location service for group members is combined with the membership management to avoid the use of separate location server. We have also handled the empty zone problem which is challenging for the zone-based protocols. We quantitatively analyze the control overhead of the proposed SOD-MRP protocol and our analysis results indicate that the per-node cost of SODMRP keeps relatively constant with respect to network size and group size. We have performed extensive simulations to evaluate the performance of SODMRP. It scales well with the group size, and achieves more than 98 % delivery ratio under all the group sizes studied. On the other hand, the delivery ratios of SPBM drop significantly when there is a large number of groups in the network or when the network size is large.

6 Result and Discussions

The NS2 network simulator is used for evaluating the performance of the protocol for multicast group communications. For all these experiments, 50 nodes are placed randomly in a 1000×1000 m^2 area and chosen a group of 20 nodes that join the multicast group. In Figs. 4 and 5 the SODMRP performance at various condition of the network is evaluated (Fig. 6).

References

1. Kopka H, Daly PW (1999) A guide to LATEX, 3rd edn. Addison-Wesley, Harlow, UK
2. Ballardie T, Francis P, Crowcroft J (1993) Core based trees (CBT)—an architecture for scalable inter-domain multicast routing. In Proceedingsof ACM SIGCOMM'93, San Francisco, CA, pp 85–95
3. Ji L, Corson MS (2001) Differential destination multicast: a MANET multicast routing protocol for small groups. In: Proceedings IEEE Infocom01 Anchorage, Alaska, April 2001
4. Royer EM, Perkins CE (1999) Multicast operation of the ad hoc on-demand distance vector routing protocol. In: Proceedings of the ACM/IEEE international conference on mobile computing and networking (MOBICOM), pp 207218
5. Wu C, Tay Y, Toh C-K (1998) Ad hoc multicast routing protocol utilizing increasing id-numbers (AMRIS) functional specification. Internet draft, November 1998
6. Zhang X, Jacob L (2003) Multicast zone routing protocol in mobile ad hoc wireless networks. In: Proceedings of local computer networks, 2003 (LCN 03), October 2003
7. Das SM, Pucha H, Hu YC () Distributed hashing for scalable multicast in wireless Ad Hoc network. In: IEEE transactions on parallel and distributed systems (TPDS), Vol 19, Issue no 3, March 2008
8. Deering SE, Cheriton DR (1990) Multicast routing in datagram internetworks and extended LANs. ACM Trans Comput Syst 8(2):85–110
9. Deering S, Estrin DL, Farinacci D, Jacobson V, Liu C-G, Wei L (1996) The PIM architecture for wide-area multicast routing. IEEE/ACM Trans Networking 4(2):153–162
10. Gerla M, Lee SJ, Su W (2000) On-demand multicast routing protocol (ODMRP) for ad hoc networks. Internet draft, draft-ietf-manet-odmrp-02.txt, 2000
11. Chiang C-C, Gerla M, Zhang L (1998) Forwarding group multicast protocol (FGMP) for multihop, mobile wireless networks. AJ, Special Issue on Mobile Computing, and ClusterComp 1(2):187–196, 1998
12. J. Yoon, M. Liu and B. Noble. Random Waypoint Considered Harmful. Proc. IEEE INFOCOM 03, 2(4), Apr. 2003
13. Ji L, Corson MS (2001) Differential destination multicast: a MANET multicast routing protocol for small groups. In: Proceedings IEEE Infocom01, Anchorage, Alaska, April 2001
14. Abraham I, Dolev D, Malkhi D (2004) LLS: a locality aware location service for mobile ad hoc networks. In workshop on discrete algorithms and methods for mobile computing and communications (DialM), 2004
15. Li J, Jannotti J, Couto DSJD, Karger DR, Morris R (2000) A scalable location service for geographic ad hoc routing. In MOBICOM, pp 120–130
16. Royer EM, Perkins CE (1999) Multicast operation of the ad hoc on demand distance vector routing protocol. In MOBICOM, pp 207–218

17. Ji L, Corson MS (1998) A lightweight adaptive multicast algorithm. In GLOBECOM, pp 1036–1042
18. Garcia-Luna-Aceves JJ, Madruga E (1999) The core-assisted mesh protocol. IEEE JSAC, pp 1380–1394

Two Stage Constrained Optimization Method to Design DC-Leakage Free Cosine Modulated Filter Banks for Image Compression

Anamika Jain and Aditya Goel

Abstract In this paper, we discuss the design of prototype filters of cosine modulated filter banks for image compression application. The design problem is formulated as a two stage nonlinear constrained optimization problem i.e. combination of residual stopband energy and coding gain. The quadratic constraints are linearized and the objective function is minimized using constrained optimization by minimizing prototype filter tap weights satisfying dc-leakage free condition. Simulation results show that the proposed design method converges within a few iterations and that high performance cosine modulated filter banks with large stopband attenuation and coding gain are obtained and results analysed on different types of images.

Keywords Coding gain (CG) · Cosine modulated filter bank (CMFB) · Peak signal to noise ratio (PSNR) · Perfect reconstruction (PR) · Prototype filter · Quadrature mirror filter (QMF)

1 Introduction

Multirate filter banks are very important in today's world, as they find application in various fields of signal, image and video processing for subband/transform coding. Filter banks are used to separate the input signal into several components,

A. Jain (✉) · A. Goel
Department of Electronics & Communication Engineering, M.A.N.I.T., Bhopal 462 001, India
e-mail: anamikajain.bpl@gmail.com

A. Goel
e-mail: Adityagoel2@rediffmail.com

Mohan S. and S. S. Kumar (eds.), *Proceedings of the Fourth International Conference on Signal and Image Processing 2012 (ICSIP 2012)*, Lecture Notes in Electrical Engineering 222,
DOI: 10.1007/978-81-322-1000-9_40,

each one carrying a single frequency subband of the original signal [1–3]. The subband filters in cosine modulated filter banks are generated by modulating a linear phase prototype filter using the cosine matrix. Thus the design and implementation complexity is very low as compared to that of ordinary M-channel filter banks. This property has made the cosine modulated filter banks a popular choice in many applications.

Recently, it has shown that cosine modulated filter banks without dc-leakage free property [4] when applied to image coding, the decoded image exhibit checker board artifacts. In this paper, we discuss the design of cosine modulated filter banks and, in particular, dc-leakage free cosine modulated filter banks for image compression. Since subband filters are modulated versions of a single prototype filter, high performance cosine modulated filter banks can be achieved by optimizing the performance of the prototype filter only. Different design strategies have been developed to design prototype filters. Koilpillai [5] derived a lattice structure that parameterizes all the filters that leads to perfect reconstruction cosine modulated filter banks using the Givens rotation angles. However, the relationship between the lattice parameters and the stopband attenuation of the prototype filter is highly nonlinear. Consequently, prototype filters that have high stopband attenuation are difficult to obtain. Nguyen has shown that the prototype filter design problem can be formulated as a least squares optimization with quadratic constraints [6]. However, the quadratic constraint is formed using nonpositive definite matrix. Therefore, it is difficult to solve the minimization problem. Although the quadratic constrained least squares problem is difficult to solve, high performance prototype filters have been obtained by numerical methods as demonstrated in [7]. Different approaches, such as the iterative method, using window method proposed in [8–12], have been adopted to modify the quadratic constraint least squares method.

In this paper, we describe an efficient modification to the quadratic constraint for the design of prototype filters to have maximum coding gain and high stopband attenuation simultaneously. Unlike, the linearization procedure in [7–9], the proposed algorithm modifies the quadratic constraint directly. As for image coding required filter bank should be smooth, linear phase and low tap to reduce computational complexity, on the other hand, for high stopband attenuation the filter taps should be high thus cost function is solved using two step optimization. Thus using a different cost function, design problem can be very flexible, and can be used to design prototype filters for dc-leakage free cosine modulated filter banks. Simulation results have shown that high performance prototype filter can be obtained by the proposed design method with a few iterations. Further, image coding results shows that designed cosine modulated filter banks performs better in spite of nonlinear characteristics of the component filters [13].

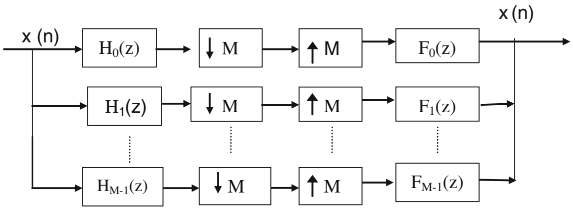


Fig. 1 M-channel maximally-decimated filter bank

2 Cosine Modulated Filter Banks

The analysis subband filters $h_k(n)$ and synthesis subband filters $f_k(n)$ as shown in Fig. 1 with $0 \leq k \leq M-1$ in a M-channel cosine modulated filter bank are given by [14]

$$h_k(n) = 2h(n)\cos\left((2k+1)\frac{\pi}{2M}\left(n - \frac{N-1}{2}\right) + (-1)^k\frac{\pi}{4}\right) \tag{1}$$

$$f_k(n) = h(N-1-n) \tag{2}$$

Where, $0 \leq n \leq N-1$ and N is the length of the linear phase prototype filter h(n). The length N is further assumed to be multiples of 2 M, *i.e.* N = 2 mM with an integer m [20, 21].

2.1 *Perfect reconstruction condition*

It has been shown in [5] that the perfect reconstruction condition for cosine modulated filter banks can be reduced to

$$G_k(z^{-1})G_k(z) + G_{M+k}(z^{-1})G_{M+k}(z) = \frac{1}{2M} \qquad 0 \leq k \leq M/2 - 1 \tag{3}$$

Where M is assumed to be even, and $G_k(z)$ are the polyphase components of H(z), such that

$$H(z) = \sum_{k=0}^{2M-1} z^{-k} G_k\left(z^{2M}\right) \tag{4}$$

To simplify our discussion, we focus on the case of M being even. The formulation for the case of M being odd will be similar. In vector matrix notation, the polyphase component $G_k(z)$ can be rewritten as

$$G_k(z) = h^t V_k e \tag{5}$$

$$G_k(z^{-1}) = z^{m-1} h^t V_k J e \tag{6}$$

$$\text{Where} \qquad h = [h(0)\ h(1)\ldots h(mM-1)]^t$$

$$e = \left[1\ \ z^{-1}\ldots z^{-(m-1)}\right]^t$$

$$V_k \in R_{mM \times m}$$

$$[V_k]_{i,j} = \begin{cases} 1 & \begin{cases} either\begin{cases} i = k + 2jM \\ k + 2jM < mM \end{cases} \\ or\begin{cases} i = 2M(m-j) - 1 - k \\ k + 2jM \geq mM \end{cases} \end{cases} \\ 0. & otherwise \end{cases}$$

The perfect reconstruction condition in (3) can be simplified as

$$h^t\big((V_k + V_{M+k})JD_n\big(V_k^t + V_{M+k}^t\big)h = \begin{cases} 0 & 0 \leq n \leq m-2 \\ \frac{1}{2M} & m = m-1 \end{cases} \tag{7}$$

Where $D_n \in R_{m \times m}$ with

$$[D_n]_{i,j} = \begin{cases} 1, & i + j = n \\ 0, & otherwise \end{cases}$$

as shown in [6]. The above condition can be written in quadratic form

$$h^t Q_{k,n} h = \beta_n \tag{8}$$

$$Q_{k,n} = (V_k + V_{M+k})JD_n(V_k^t + V_{M+k}^t) \tag{9}$$

$$\beta_n = \begin{cases} 0 & 0 \leq n \leq m-2 \\ \frac{1}{2M} & m = m-1 \end{cases} \tag{10}$$

Apart from perfect reconstruction, the performance of cosine modulated filter banks is also characterized by the stopband attenuation of the subband filters.

2.2 Stopband Attenuation

The stopband attenuation of the subband filters are directly related to the stopband attenuation of the prototype filter. As a result, the stopband attenuation of the prototype filter is used to form the cost function in the design problem. The magnitude response A (ω) and the stopband energy of the linear phase prototype filter H (z) can be computed as

$$E_s = \int_{\omega_s}^{\pi} A^2(\omega)d\omega \tag{11}$$

$$A(\omega) = h^t c(\omega) \tag{12}$$

$$c(\omega) = \left[\cos\frac{(mM-1)\omega}{2}\ \cos\frac{(mM-2)\omega}{2} \ldots \cos\frac{\omega}{2}\right]^t \tag{13}$$

Where $\omega_s = \pi/(2\ \mathrm{M}) + \in$ is the stopband edge of the prototype filter. Note that $\in$, a small positive number, defines the transition width of the prototype filter. The above integration can be approximated by a finite sum of the A (ω) with ω being the samples in the stopband region.

$$E_s \approx 4h^t Ph \tag{14}$$

$$P = CC^t \tag{15}$$

$$C = [c(\omega_0)\ c(\omega_1)\ldots c(\omega_{L-1})] \tag{16}$$

$$\omega_k = \frac{\pi}{2M} + \frac{(2M-1)(2k+1)\pi}{4ML} \tag{17}$$

Where, L is the number of samples in the stopband region.

2.3 DC-Leakage Free Condition

It has been found that dc-leakage free filter banks are important in image coding applications as shown in [7] that processing images with a non-dc-leakage free filter bank results in checkerboard effects when the subband coefficients are quantized. A dc-leakage free filter bank has all the subband filters, except the lowpass subband filter, has at least one zero at z = 1. As a result, all the bandpass and highpass subband filters are dc transmission free or are free from aliasing errors. Since the subband filters in cosine modulated filter banks are the modulated version of the prototype filter, the cosine modulated filter is dc-leakage free if and only if

$$h^t c(\omega) = 0, \quad \forall\omega = \frac{(2k+1)\pi}{2M}, 1 \le k \le M-1 \tag{18}$$

2.4 Coding Gain

In image compression, higher coding gain correlates most consistently with higher objective performance (PSNR). Transforms with higher coding gain compact more energy into a fewer number of coefficients. All design examples in this paper are obtained with a version of the generalized coding gain formula given by:

$$E_{codinggain} = 10\log_{10}\frac{\sigma_x^2}{\left(\prod_{k=0}^{M-1}\sigma_{x_i}^2\|f_i\|^2\right)^{1/M}} \tag{19}$$

$$\sigma_{x_i}^2 = h_i^t R_{xx} h_i \tag{20}$$

Where is variance of the input signal; $\sigma_{\mathrm{x}_i}^2$ is variance of the i$^{\mathrm{th}}$ subband; $\|\mathrm{f}_i\|^2$ is norm of the i$^{\mathrm{th}}$ synthesis filter. The signal is the commonly-used AR (1) process with intersample autocorrelation coefficient $\rho = .95$ [15].

Problem formulation

Combined with the perfect reconstruction and dc leakage free condition condition, the design problem can be expressed as quadratic constrained optimization problem of the prototype filter coefficients in the following form:

$$\begin{aligned} \min_n (i)\alpha E_s \\ (ii)\beta E_{codinggain} \end{aligned} \tag{21}$$

$$\begin{aligned} \textit{Subject to} \quad & (a)\ h^t Q_{k,n} h = \beta_n \\ & (b)\ h^t C(\omega) = 0, \end{aligned}$$

$$\forall\omega = \frac{(2k+1)\pi}{2M}, 1 \le k \le M-1 \quad \text{With}$$

$$\begin{cases} 0 \le k \le \frac{M}{2} - 1 \\ 0 \le n \le, m-1 \end{cases}$$

Where P is positive definite. Hence, both the cost function in the above minimization problem is highly nonlinear thus two stage nonlinear constrained optimization methods [17] is used to solve the above problem.

3 The Design Algorithm

In the algorithm presented here, the unit energy constraint [16] is imposed within some pre specified limit. The design algorithm proceeds through the following steps:

(1) Assume initial values of α and β, δ and ω_{s}
(2) Start with an initial vector $h_0 = [h(0), \ldots, h(N)]^t$ satisfying the unit energy constraint.
(3) Set the function tolerance, convergence tolerance $p = \left[h\left(\frac{N+1}{2}\right), \ldots, h(N)\right]^t$

(4) Optimize objective function eqn. 21 (i) using constrained optimization method for the smallest objective value for the constraints satisfied within specified limit.
(5) Evaluate all the analysis and synthesis component filters of cosine modulated filter bank using this prototype filter h.
(6) Now, taking the designed filter bank as the starting point optimize the second objective function eqn. 21(ii) *i.e.* maximum coding gain satisfying the constraints again.
(7) Repeat step (4)-(6) until both the criterion satisfied within some tolerance.
(8) Evaluate all other performance parameters for prototype and component filters of the overall filter bank.

4 Results and Discussion

4.1 Design Examples

Using the proposed approaches, we designed 8-channel 16 taps (8x16) and 8-channel 24 taps (8×24) cosine modulated filter bank with and without dc leakage giving highest priority to coding gain. The magnitude response of the designed dc leakage free filter bank (8×16) in Fig. 2 a and 2b indicate that stopband attenuation is (15 dB) much lower as compared to filter bank with dc leakage i.e. (22 dB). Figure 3 a and 3b indicates increase in stopband attenuation with increasing no. of taps i.e. (19 dB), as compared to the (8×16) dc leakage filter bank but lower i.e. (32 dB) as compared to filter bank with dc leakage. The dc leakage free condition is achieved at the cost of reduced stopband attenuation as the perfect reconstruction constraints were already added to the design problem. It can be seen that dc leakage free condition improve the performance of subband filters and reduce the aliasing errors as all the subband filters except the lowpass one, has at least one zero at $z = 1$ and thus all other bandpass and highpass filters are dc transmission free [7].

The filter bank, designed to maximize coding gain, without dc leakage have smooth response to eliminate blocking artifacts completely so that the resulting reconstructed images are visually pleasant. As compared in Table 1, 8×16 CMFB in design example I, with dc leakage (optimized for pure coding gain), attain 9.3749 dB, which almost equals the coding gain of optimal biorthogonal systems. However, applying dc leakage free constraints on the filter coefficients ensure a high level of perceptual performance in image coding. In design example II, the 8×24 case, our CMFB achieves an improvement of 0.001 dB in coding gain. CMFB's with filter length 16 and 24 are shown in Figs. 2 and 3. Increasing the length only improves the coding gain marginally, it helps in the case of

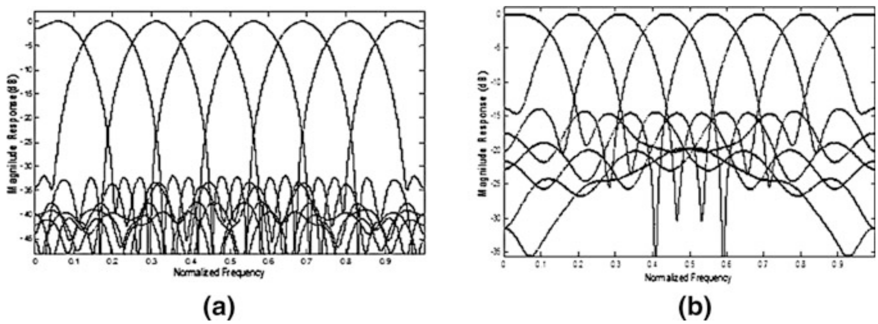


Fig. 2 Magnitude response of the 8 × 16 Cosine modulated filter bank, (**a**) with dc leakage, (**b**) without dc leakage

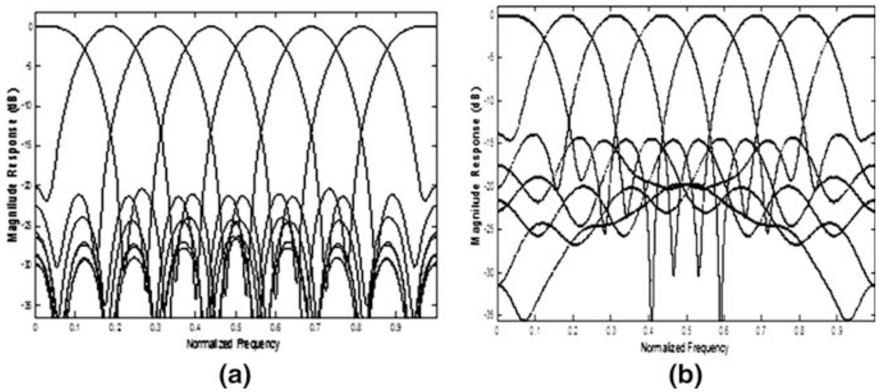


Fig. 3 Magnitude response of the 8 × 24 Cosine modulated filter bank (**a**) with dc leakage, (**b**) without dc leakage

stopband attenuation (where longer filters are always beneficial), which result in more complexity in design and implementation.

4.2 Application in Image Coding

The performance of our designed Cosine Modulated Filter Bank's is evaluated through an image coding comparison among the DCT and CMFB with and without dc leakage. In all simulations, we employed a single level decomposition for the 8-channel transforms. The coder algorithm used is the JPEG baseline system [15], merely replacing the DCT by our designed filter bank obtained through separable implementation of the 1-D transform. Instead of the symmetric extension, we used the extension method in [13] for our designed CMFB's.

Table 1 Comparison of different transform's performance

Transform	Coding gain	Stopband attenuation(dB)
8 × 8 DCT	8.83	9.96
8 × 16 LOT	9.22	19.38
8 × 24 GenLOT	9.35	23.20
Proposed (8 × 24)	9.3749	26.026

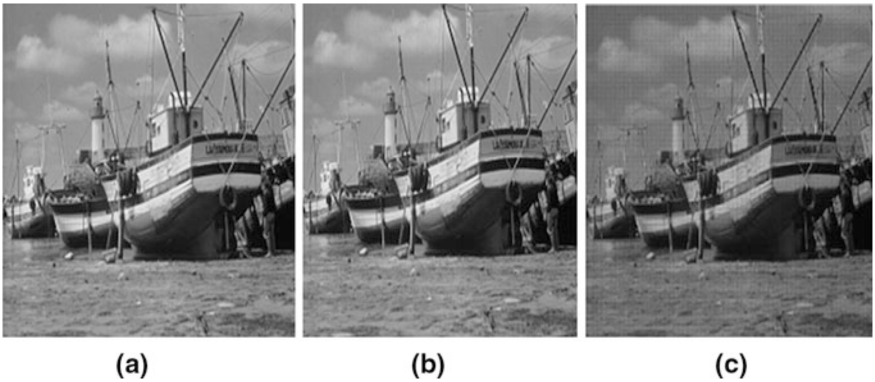


Fig. 4 **a** Original 'Boat' image. **b** Reconstructed image without dc leakage. **c**Reconstructed image with dc leakage

The images chosen for the coding experiments are Barbara and Boat. All of them are standard, well known 512 × 512 8-bit gray-scale test images. The objective distortion measure is the popular peak signal-to-noise ratio (PSNR)

$$PSNR = 10\log_{10}\left(\frac{255}{MSE}\right)dB \tag{22}$$

Where, MSE denotes the mean squared error.

Two versions of filter bank one is without dc leakage and the other is with dc leakage were included in the tests, and both were optimized for maximum coding gain. The reason for the inclusion of the two types of filter banks is that with dc leakage, one can achieve higher coding gain and perhaps, higher peak signal-to-noise ratio (PSNR) after decompressing the image. However, the design without dc leakage yields decompressed images with higher visual quality. Reconstructed versions of image "Barbara" and "Boat" using designed filter bank (8 × 16) with and without dc leakage are shown in Figs. 4, 5.

For a image like Boat, performance of the 8 × 24 dc leakage free CMFB is shown in Fig. 4b and c. The images without dc leakage filter bank produce are superior as in low bit-rate coding image is reconstructed using only nonzero coefficients which are few in number and are very concentrated in lowest frequency component. Thus without dc leakage lowpass filter output lead to a smooth output and reconstructed image quality is superior as details are preserved as

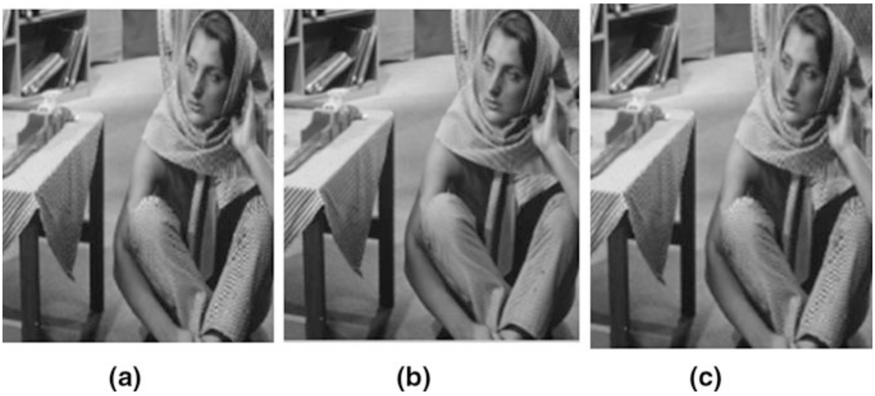


Fig. 5 **a** Original 'barbara' image. **b** Reconstructed image without dc leakage. **c** Reconstructed image with dc leakage

Table 2 Image coding results with and without dc leakage barbara' image (0.25 bpp)

M	Length	With dc leakage		Without dc leakage	
		PSNR	CG	PSNR	CG
8	16	28.02	8.89	29.99	8.85
8	24	30.11	9.37	31.18	9.21
8	32	32.02	9.374	33.63	9.257

shown in Fig. 4c even though coding gain is lower. For a highly textured image like Barbara, performance of the 8×24 dc leakage frees CMFB shown in Fig. 5 b and c. Results for 8 channel filter bank with different lengths as tabulated in Table 2 for 0.25 bpp shows that with dc leakage filter bank a PSNR gain of around 2.5 dB is obtained over a wide range of bit rates. Reconstructed image quality without dc leakage filter bank, is superior as texture is preserved, blocking is eliminated, and ringing is barely noticeable as shown in Fig. 5c.

5 Conclusion

The design of prototype filters for cosine modulated filter banks is discussed in this paper. The design problem is solved in two stages as a constrained optimization problem as in [17] to design two channel filter banks in subband image coding. We use linearize quadratic constraint as proposed in [18], since the constraint matrix is not positive definite, Design examples were presented for both cosine modulated filter banks with and without dc-leakage. Performance of filter banks is evaluated in terms of both PSNR and visual quality on different type of images. The work can be further extended to design of biorthogonal cosine modulated filter banks for image coding application.

References

1. Esteban D, Galand C (1997) Application of quadrature mirror filter to split band voice coding schemes. In: proceedings of IEEE international conference on acoustics, speech, and signal processing (ASSP) pp 191–195
2. Smith MJT, Eddins SL (1990) Analysis/synthesis techniques for sub-band image coding. IEEE Trans Acoust Speech Signal Process ASSP 38(8):1446–1456
3. Woods JW, O'Neil SD (1986) Sub-band coding of images. IEEE Trans Acoust Speech Signal Process ASSP 34(10):1278–1288
4. Karp T, Mertins A (1998) Biorthogonal cosine-modulated filter banks without DC leakage. Proc IEEE ICASSP :1457–1460
5. Koilpillai RD, Vaidyanathan PP (1992) Cosine-modulated FIR filterbanks satisfying perfect reconstruction. IEEE Trans Signal Process 40:770–783
6. Nguyen TQ (1995) Digital filter bank design quadratic constrained formulation. IEEE Trans Signal Process 43:2103–2108
7. Yeun-Ting F, Chi-Wah K (2003) Iterative least squares design of dc-leakage free paraunitary cosine modulated filter banks. IEEE trans circuits syst—II: analog digital signal process 50(5):238–243
8. Xu H, Lu WS, Antoniou A (1996) Efficient iterative design method for cosine-modulated QMF banks. IEEE Trans Signal Process 44:1657–1668
9. Abdel-Raheem E, El-Guibaly F, Antoniou A (1995) Efficient design of cosine-modulated filter banks. In: proceedings of IEEE pacific rim conf.on communications, Computers, and Signal Processing, pp.387–390
10. Saramaki T, Bregovic R (2001) An efficient approach for designing nearly perfect reconstruction cosine-modulated and modified DFT filter banks. Proc IEEE ICASSP 6:3617–3620
11. Creusere Charles D, Mitra Sanjit K (1995) A simple method for designing high-quality prototype filters for m band pseudo QMF banks. IEEE Trans Signal Process 43:1005–1007
12. Lin YP, Vaidyanathan PP (1998) A Kaiser window approach for the design of prototype filters of cosine modulated filter banks. IEEE Signal Process Lett 5:132–134
13. Uto T, Oka T, Ikehara M (2007) M-channel nonlinear phase filter banks in image compression: Structure, design, and signal extension. IEEE Trans. Signal Process 55:1339–1351
14. Heller PN, Karp T, Nguyen TQ (1999) A general formulation of modulated filter banks. IEEE Trans Signal Process 47:986–1002
15. Katto J, Yasuda Y (1991) Performance evaluation of subband coding and optimization of its filter coefficients. In: Proceedings of SPIE visual communication and image processing Nov pp 95–106
16. Jain VK, Crochiere RE (1984) Quadrature mirror filter design in time domain. IEEE Trans Acoust. Speech Signal Process ASSP 32(4):353–361
17. Wah Constrained optimization of filter bank for subband image coding
18. Nguyen TQ, Heller PN (1996) Biorthogonal cosine-modulated filter bank. Proc IEEE ICASSP 3:1471–1474
19. De Queiroz RL, Nguyen TQ, Rao KR (1996) The genlot: generalized linear-phase lapped orthogonal transform. IEEE trans signal process 44(3):497–507
20. Vaidyanathan PP (1993) Multirate systems and filter banks. Prentice Hall, Englewood Cliffs, NJ
21. Strang G, Nguyen TQ (1996) Wavelets and Filter Banks. Wellesley Cambridge, Wellesley
22. Gao X, Bo H, Luo L, He Z (1997) Fractal image coding using cosine-modulated filter banks. IEEE Trans Consumer Electron 43(1):62–68
23. Bolecski H, Stranz T, Hlawatsch F, Sucher R(1997) Subband image coding using cosine modulated filter bank with perfect reconstruction and linear phase. In: proceedings of IEEE ICIP-97, October, pp 594–597

Evaluation of Defect Detection in Textile Images Using Gabor Wavelet Based Independent Component Analysis and Vector Quantized Principal Component Analysis

S. Anitha and V. Radha

Abstract Textile defect detection plays an important role in the manufacturing industry to maintain the quality of the end product. Wavelet transform is more suitable for quality inspection due to its multi-resolution representation. The Gabor Wavelet Network provides an effective way to analyze the input images and to extract the texture features. The paper addresses the functionality of Gabor wavelet network with independent component analysis and vector quantized principal component analysis. The two methods are used to extract the features from the template image. Then the difference between the template image and the input image features are compared, and threshold value is calculated using Otsu method to obtain the binary image. The performances of the methods are evaluated to verify the efficiency in identifying the defect in the pattern fabric image.

Keywords Defect detection · Gabor wavelet · Fabric · Independent component analysis · Principal component analysis

1 Introduction

The automated visual inspection has been increasingly popular among many industries like steel, ceramics, wood, wallpaper etc. Fabric defect detection is one of the most important phases in fabric production to improve the fabric quality.

S. Anitha (✉) · V. Radha (✉)
Department of Computer Science, Avinashilingam University, Coimbatore, TamilNadu, India
e-mail: mail.anitha08@gmail.com

V. Radha
e-mail: radharesearch@gmail.com

Mohan S. and S. S. Kumar (eds.), *Proceedings of the Fourth International Conference on Signal and Image Processing 2012 (ICSIP 2012)*, Lecture Notes in Electrical Engineering 222,
DOI: 10.1007/978-81-322-1000-9_41,

Generally fabric inspections are done by human experts and inspectors which involve high cost and low performance. While using manual inspection only 80 % of fabric defects can be identified [1]. For the past decade considerable research has been carried out to automate the fabric inspection and the ultimate objective of the research is to find efficient approach to increase accuracy meanwhile to reduce complexity and cost.

A fabric defect is any abnormality in the fabric that hinders its acceptability by the consumer. The textile processing does not eliminate variability incurred during different steps in textile manufacturing. As materials flow from one stage of processing to another, components of variability are added and the final product involves a cumulative variability that is much higher than the variability of the input fibers and thus it cause a defect in the fabric. The main factors that lead to fabric defects are failure of opening and cleaning the machines that completely eliminate contaminants and trash particles, and it may leads to spinning, weaving and knitting related defects. So the fabric inspection has to identify all types of defects with minimum effort.

Textile fabric can be broadly categorized into two types that include plain and patterned fabric. Considerable work has been carried out in the plain fabrics and promising approaches are evolved. However, very few researchers are involved in patterned fabric. The 'patterned' fabric is defined as fabric with repetitive patterned units on its design. Under the patterned fabric, there are many categories where the repetitive unit can range from simplest dots, checked box, stripes to the most complicated pattern like flower, animals or designed patterns like wall paper groups and it is depicted in Fig. 1 classification of fabric [2].

Several methods have been proposed to address the problem of detecting defects in textile fabrics, including statistical, spectral and model based approaches. The wavelet transform provides a solid and unified mathematical framework for the analysis and characterization of an image at different scales. It provides both time and frequency information, and can be successfully applied for textile defect detection. Fabric defect detection based on wavelet transform has performs better with less computation than the traditional statistical texture analysis approaches in identifying defects. A Gabor filter has an optimal localization both in the spatial domain and in the spatial frequency domain and it is one of the most famous spectral approaches that are widely used in the field of defect detection.

Independent component analysis (ICA) is essentially a method of extracting individual signals from the mixture of signals. ICA is a computational method [3] for separating a multivariate signal into additive subcomponents that are statistically independent of the non-Gaussian source signals. Independent component analysis method is combined along with wavelet transforms for identifying defects in textile fabric images [4]. Principal Component Analysis (PCA) is also referred as Eigen face decomposition [5] and is a standard, popular approach used in pattern recognition and signals processing, since it is a simple and non-parametric method [6] for data extraction and reduction, and performs better in the field of texture analysis [7]. As the pattern often contains redundant information, mapping it to a feature vector can decrease the redundancy and it preserve most of the major

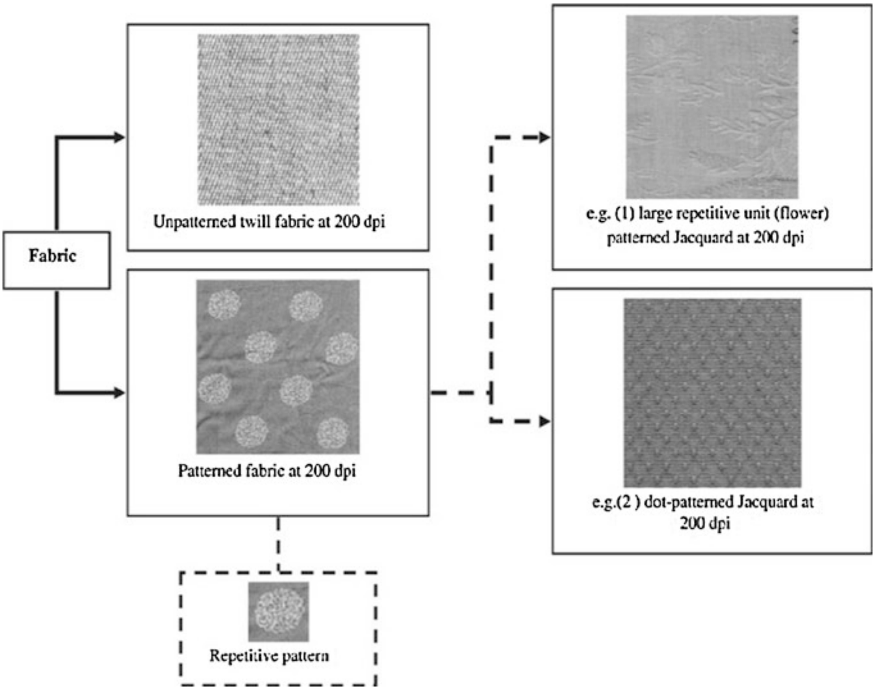


Fig. 1 Classification of fabric

information content of the pattern that are called principal components and is suitable to be applied in the field of defect detection. The Gabor wavelet network is combined with both ICA and Vector Quantized (VQ) PCA for better performance.

In defect detection process 90 % of the defects in a plain fabric could be detected simply by applying a threshold method [8]. The traditional methods for plain fabric inspection cannot be applied on patterned fabric inspection. Traditional methods involve complexity to discriminate between the pattern and the defect, hence the feature extraction requires more effort. Only a few research studies are found to be related to patterned fabric inspection and are mostly evaluated with the dot patterned Jacquard.

In this paper Gabor wavelet network is combined with two different approaches, ICA and VQ PCA and the performance of these two methods are evaluated on patterned fabric especially for Jacquard patterned fabric. The Gabor wavelet network is given in Sect. 2. Section 3 outlines the ICA method. Then, VQ PCA method is presented in Sect. 4. Experimental results and discussion are presented in Sect. 5. Finally, conclusion is presented in Sect. 6.

2 Gabor Wavelet Network

Gabor filters have been successfully implemented in various approaches of image analysis and computer vision applications. Gabor filters can effectively combine both spatial and frequency domain textile information, hence it is suitable for defect detection application. Generally GWN is a combination of Feed Forward Neural Network (FFN), namely Multi Layer Perceptron (MLP) and the Gabor wavelet decomposition. Various experiments [9], [1, 10] show that GWN is an effective and task-specific feature extractor.

GWNs represent an object as a linear combination of Gabor wavelets and the parameters of each single Gabor functions (such as orientation, position and scale) are individually optimized to reflect the particular local image structure. Gabor Wavelet Networks have several advantages:

(1) GWN allows an efficient and sparse coding while coding is adaptive to the task at hand.
(2) Gabor filters are good feature detectors [1] and the optimized parameters of each of the Gabor wavelets are directly related to the underlying image structure.
(3) The wavelet coefficients (or weights) of each of the Gabor wavelets are linearly related to the filter responses and with that they are also directly related to the image structure.
(4) The precision of the representation can be varied to any desired degree ranging from a coarse representation to an almost photo-realistic one by simply varying the number of used wavelets.
(5) GWNs are invariant to affine deformations without shear and homogeneous illumination changes [11], [2].

GWN allows the tuning of filter parameters to match a particular texture feature, such as orientation and central frequency. GWN is used to extract image feature which can be used in pattern recognition. The prior information for the design of optimal Gabor filter is obtained from GWN [12], [9]. The GWN with single hidden layer can extract local features form non-defective texture pattern. However, single layer in GWN, requires the initial parameters of wavelet to be selected carefully.

The wavelet networks proposed [10] the concept of Gabor Wavelet for solving the 2D problems in pattern recognition [13], in which an imaginary Gabor wavelet function is used as a transfer function in the hidden layer of the network. The mapping form of the network can be governed by Eq. (1).

$$f(x,y) = \sum_{i=1}^{N} w_i g_0^i(x,y) + \bar{f} \tag{1}$$

where w_i is a network weight from the hidden layer to the output layer and $\bar{f}$ is introduced to eliminate the DC value of an objective function. The imaginary part of the Gabor function is used and it is referred as the transfer function, and is expressed in Eq. (2).

$$g_{o=}^{i}\exp\left\{-\frac{\left[\left(x-t_x^i\right)\cos\theta^i-\left(y-t_y^i\right)\sin\theta^i\right]^2}{2\left(\sigma_x^i\right)^2}-\frac{\left[\left(x-t_x^i\right)\sin\theta^i-\left(y-t_y^i\right)\cos\theta^i\right]^2}{2\left(\sigma_y^i\right)^2}\right\}$$
$$\times\ \sin\left(2\pi\omega_{\mathrm{x}}^{\mathrm{i}}\left[\left(\mathrm{x}-\mathrm{t}_{\mathrm{x}}^{\mathrm{i}}\right)\cos\theta^{\mathrm{i}}-\left(\mathrm{y}-\mathrm{t}_{\mathrm{y}}^{\mathrm{i}}\right)\sin\theta^{\mathrm{i}}\right]\right) \tag{2}$$

where t_x^i, t_y^i are the translation parameters of the i^{th} Gabor wavelet, and $\left(\sigma_x^i, \sigma_y^i\right)$, θ^i and ω_x^i are the radial frequency bandwidths, the orientation and the central frequency respectively of the i^{th} hidden node. The network input vector [x, y] is the position of a pixel in a studied image IM, and the output is the grey level of the corresponding pixel.

Figure 2 depicts the architecture for a Gabor wavelet network. In the network, there are five parameters for each Gabor wavelet, which should be determined by the network learning process, including the translation parameters, orientation, radial frequency bandwidth, centre frequency, and the corresponding weight. The objective function of the learning process is defined as given in Eq. (3).

$$\mathrm{E}=\min\left\|\mathrm{IM}-\sum\nolimits_{\mathrm{i}} w_i g_i\right\|_2^2 \tag{3}$$

In fact, the network proposed in [9] has only two input nodes and one output node. In the network, the input vector [x y] is the position of a pixel in the template image, and the output is the grey level of the corresponding pixel. The GWN is offered with a supervised training with the non defective fabric image as the template and it is used to determine the parameters of optimal Gabor filter. Once

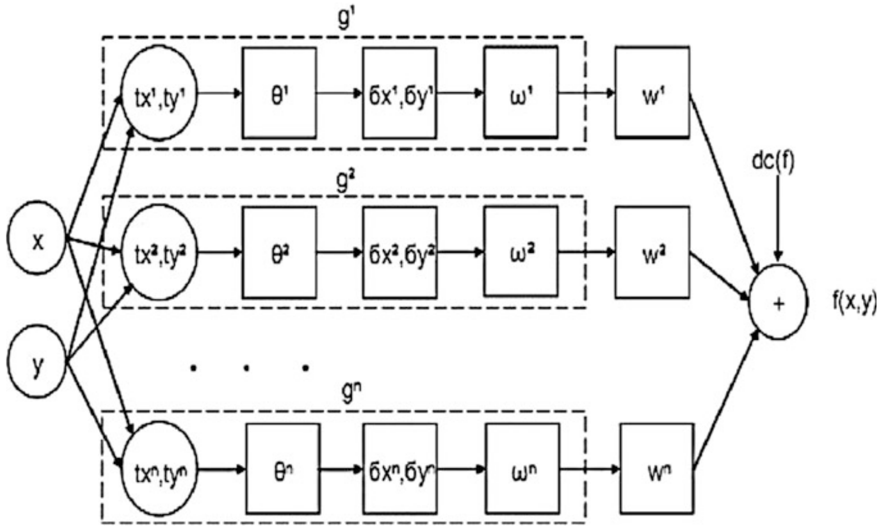


Fig. 2 The structure of a gabor wavelet network

the network is trained, the optimal filter is used to discriminate defective and non defective from the fabric images with the same texture background as in the template image.

3 Independent Component Analysis

Independent component analysis (ICA) is fundamentally a method for extracting individual signals from mixtures of signals [14]. The goal of ICA is to linearly transform the data such that the transformed variables are as statistically independent from each other as possible [8]. ICA has proven a useful tool for finding structure in data. ICA separates the different independent components (ICs) of a signal without making use of any specific knowledge of the component signals, and it is referred as blind source separation [15, 16].

The simple and generic nature of the assumption ensures that ICA is successfully applied to a diverse range of research fields especially in pattern recognition. ICA basis functions capture the inherent properties of a texture. ICA based defect detection use defect free texture images as the training data and it is used to identify the input image as defective or non defective [17,18].

ICA extracts statistically independent components, S, from a set of measured signals, X, as shown in Eq. (4).

$$X = AS \tag{4}$$

The goal of ICA is to estimate the mixing matrix A, and the source vector S, from the measured data X. ICA assumes that the different sources are independent and based on this assumption estimation is made using the mixing matrix recursively. The Gabor wavelet combined with ICA involves two steps. In the first step, a wavelet analysis is applied for each channel and each channel is split into a number of wavelet components with different scale and orientation.

The extracted wavelet features are used as input for the ICA and it identifies the independent sources. Thus the independent features are extracted from the defect free image and averaging is performed to calculate ICA basis vectors. Similarly the same process is applied for the input image and the ICA basis vector is calculated. The basis vectors of the template image is compared with the input image and based on the difference, the input image is labeled as either defective or non defective by applying a threshold technique.

4 Vector Quantized Principal Component Analysis

Generally PCA is used to find a low-dimensional representation model for high dimensional data and it is used for dimensionality reduction. Principal Component Analysis is a popular technique for data compression and has been successfully used as initial step in many computer vision tasks [2] [19]. The principal components of a set of process variables 1x, 2x, …, px are just a particular set of linear combinations of these variables. Geometrically, the principal component variables 1y, 2y, …, py are the axes of a new coordinate system obtained by rotating the axes of the original system.

The new axes represent the directions of maximum variability. The basic intent of principal components is to find the new set of orthogonal directions that define the maximum variability in the original data which is lesser than the original p variables. The information contained in the complete set of all p principal components is exactly equivalent to the information in the complete set of all original process variables, thus it helps to obtain a satisfactory description [3].

Because of its ability to discriminate directions with the largest variance in a data set, PCA is suitable to identify the most representative features as inputs to defect detection. Since it is difficult to identify patterns in high dimensional data, PCA is a powerful tool for analyzing data by reducing the number of dimensions, without much loss of information. PCA encodes textural information only, while geometrical information is discarded. Vector Quantization considers the low level information from the image block, hence it is suitable to produce accurate results. The VQPCA is applied to the wavelet coefficients obtained from GWN and provides the components of feature vectors for the template image. The same process is repeated for the input image. The difference between the input image and the template image is compared, based on the threshold value, the input image is labeled either as defective or non defective.

Then vector quantized principal component analysis which is generally used for image representation purposes are considered the defect detection task. Generally the techniques achieve high detection accuracy but the requirement of data storage and processing time are expensive.

5 Experimental Results and Discussion

The performance of the two methods GWN with ICA and GWN with VQPCA is evaluated using thirty test images. From the 30 test images 15 were defect free and 15 were defective images. The test images of textile fabric samples are of 256×256 pixels in size. The Gabor wavelets were projected by using four scales and six directions.

The important step during fault detection is the removal of noise. A common type of noise that affects patterned fabric images is the impulse noise. So Enhanced Directional Switching Median Filter (EDSMF) works better in removing impulse noise with low and high noise density while preserving important image and edge details of fabric image [20]. The EDSMF significantly increases the reliability of defect detection.

In the first method GWN is trained with the ICA of the defect free image. The features are extracted from the defect free images and are compared with the input

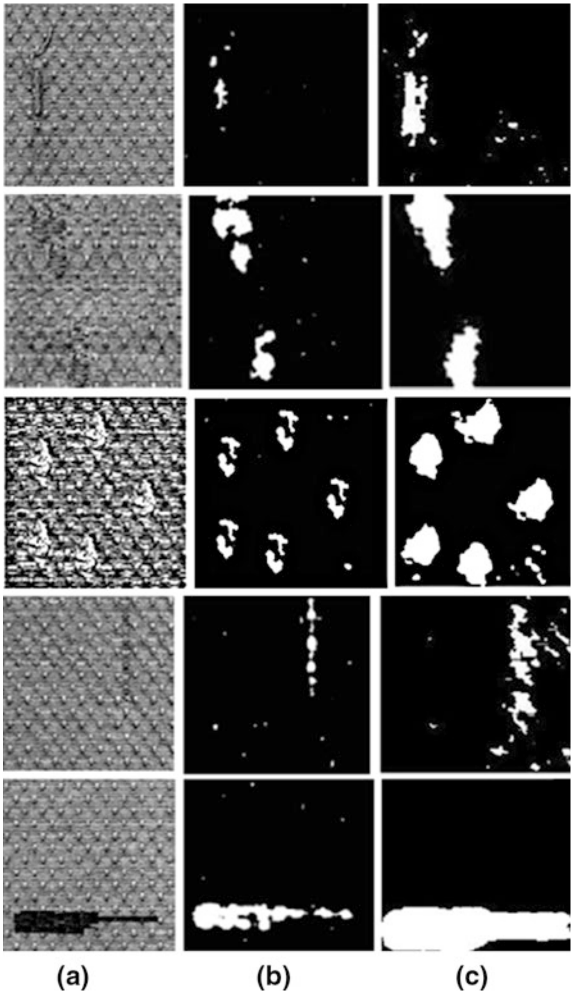


Fig. 3 Various defective sample and binary segmented results **a** input image, **b** ICA based method, **c** VQPCA based method

image, it uses the Otsu thresholding method for discrimination and segmentation is performed to obtain the binary image. In the second method GWN is used with VQPCA. VQPCA is used to find the pattern in the high dimensional image data and it is used to reduce the number of dimensions, without much loss of information. The features obtained from template image by applying GWN with VQPCA is compared with the features of input image. And thresholding is applied to generate the binary image as the output.

The performances of these methods are evaluated based on the success rate of identifying both defect and non defective image. Some of the randomly selected defective fabric images and the results are obtained by applying GWN with ICA and GWN with VQPCA are given in Fig. 3. Fourteen out of 15 defect-free images

Table 1 Comparison of detection success rate for the two main methods

30 Test samples	Comparison of two methods	
	GWN + ICA	GWN + VQPCA
15 Defect-free images	13	14
15 Defective images	12	13
Overall detection Success rate	83.3 %	89.99 %

and thirteen out of 15 defective patterned images could be correctly detected using GWN based VQPCA.

From the above results given in Table 1 compares the detection rate of two methods and the success in detection rates are 93.33 % for defect free images and 86.66 % for defective patterned images. The overall success detection rate of GWN based VQPCA is 89.99 %.Both from the segmented and success rate results it is evident that the GWN based VQPCA better discriminates the defect from the background.

6 Conclusion

In this paper Gabor wavelet based methods are used to identify the defect in patterned textile images. The Gabor wavelet is a popular method for feature extraction because of its multi-resolution and multi-orientation property. The Gabor wavelet is combined with ICA and VQPCA. The performance of the Gabor wavelet with ICA and VQPCA are evaluated. The functionality of the approaches are evaluated based on the segmented results of the input image and success rate. Based on the experimental results, it is evident that the VQPCA mechanism out performs the ICA based method. So Gabor wavelet combined with VQPCA performs well in identifying the defect from the dot patterned fabric.

References

1. Manjunath B, Chellappa R (1993) A unified approach to boundary perception: edges, textures, and illusory contours. IEEE Trans Neural Netw 4(1):96–107
2. Kr̆uger V (2001) Gabor wavelet networks for object representation, Tech. Rep CS-TR-4245, University of Maryland, CFAR, May
3. Mak KL, Peng P, (2005) "Defect detection in textile fabrics using gabor networks",18th international conference on computer applications in industry and engineering, USA, Nov 9–11, pp 226–231
4. Serdaroglu A, Ertuzun A, Ercil A (2006) Defect detection in textile fabric im-ages using wavelet transforms and independent component analysis. Pattern Recog Image Anal 16(1):61–64
5. Turk M, Pentland A (1991) Eigenfaces for recognition. J Cognitive Neurosci 3(1):71–86

6. Shlens (2009) “A tutorial on principal component analysis.” {http://www.snl.salk.edu/$\sim$shlens/
7. Eleyan A, Demirel H (2007) “PCA and LDA based neural networks for human face recognition recognition”. In: Delac K, Grgic M (eds) Face recognition, June
8. Wang J, Campbell RA, Harwood RJ (1995) “Automated inspection of car-pets”. In: Proceedings of SPIE, vol 2345. pp 180–191
9. Krueger, Feris R (2001) “Wavelet subspace method for real- time face tracking”, Pattern Recognition, 23rd DAGM Symposium
10. Zhang Q, Benveniste A (1992) Wavelet networks. IEEE Trans Neural Netw 3(6):889–898
11. Daubechies I (1990) The wavelet transform, time-frequency localization and signal analysis. IEEE Trans. Inf Theory 36:961–1005
12. Mak KL, Peng p, Lau HYK (2005) ”Optimal morphological filter design for fabric defect detection”, IEEE international conference on industrial technology, China,14–17 Dec, pp 799–804
13. Krueger V, Sommer G (2000) Gabor wavelet networks for object representation. DAGM Symposium, Germany, Sept, pp 13–15
14. Hyvärinen A, Oja E (2000) Independent component analysis: algorithms and applications. Neural Network. 13(4/5):411–430
15. Bell A, Sejnowski T (1995) An information-maximization approach to blind separation and blind deconvolution. Neural Comput 7(6):1129–1159
16. Cardoso J, Cnrs P (1998) Blind signal separation: statistical principles. Proc IEEE 86(10):2009–2025
17. Sezer O, Ertüzün A, Erçil A (2004) Independent component analysis for texture defect detection. Pattern Recognit Image Anal 14(2):303–307
18. Tsai D, Lai S (2008) Defect detection in periodically patterned surfaces using independent component analysis. Pattern Recogn 41(9):2812–2832
19. Rogerio S, F Roberto M (2000) “Tracking facial features using Gabor wavelet networks”, Computer graphics and image proceeding. In: Symposium on Brazalian pro-ceedings XII, pp 22–27
20. Anitha S, Radha V (2012) “Enhanced switching median filter for denoising in 2d patterned textile images”. In: Proceedings of ICMOC, vol 38. Pp 3362–3372
21. Kruger V, Sommer G (2002) Gabor wavelet networks for efficient heard pose estimation. Image Vis Comput 20:665–672

Noise Reduction in Urdu Document Image–Spatial and Frequency Domain Approaches

R. J. Ramteke and Imran Khan Pathan

Abstract With advancement in optical character recognition technology, now it is possible to digitize printed and handwritten documents and to make it editable and searchable for many scripts and languages. But still the major challenges which need to be simplify in case of Urdu script is segmentation dilemma. The segmentation of Urdu text is untouched by most of the researchers due to complexity in Urdu script. An ideal preprocessing for Urdu script may reduce these complexities and simplify the segmentation process. The noise removal in Urdu is complex due to importance of dots and modifiers which are similar to noise. In character recognition system preprocessing intends to remove/reduce the noise, normalize image against present variations like skewness, slant, size etc. and minimize the storage requirement to increase processing speed. In present paper an attempt is made to recapitulate various preprocessing techniques proposed in literature for Arabic, Persian, Jawi and Urdu. Also the enhancement of the dark and noisy Urdu document is done using histogram equalization, spatial max and median filter, and frequency domain Gaussian Lowpass Filters. These noise free document image can help to improve further segmentation and feature extraction process.

Keywords Noise reduction · Histogram equalization · Spatial filter · Max filter · Median filter · Frequency domain gaussian lowpass filters · Normalization · Slant and skew correction

R. J. Ramteke (✉) · I. K. Pathan
School of Computer Sciences, North Maharashtra University, Jalgaon, Maharashtra, India
e-mail: rakeshj.ramteke@gmail.com

I. K. Pathan
e-mail: imk.pathan@gmail.com

Mohan S. and S. S. Kumar (eds.), *Proceedings of the Fourth International Conference on Signal and Image Processing 2012 (ICSIP 2012)*, Lecture Notes in Electrical Engineering 222,
DOI: 10.1007/978-81-322-1000-9_42,

1 Introduction

With rapid development in printing technologies and its affordable access to the people, the volume of printed material is boosting with every passing day. With increase of this printed material several challenges are encountered like searching the required information from this huge printed material and reuse the printed text by converting it into editable form. After considerable progress in character recognition now it's possible to edit or search document images; but still for many languages there is lack of robust character recognition system. Urdu is one these languages which fall behind in this race and need more efforts to resolve the dilemma of character segmentation and recognition. The major challenges which need to be simplify in case of Urdu character recognition is segmentation. The segmentation of Urdu text is untouched by most of the researchers due to complexity in Urdu script. An ideal preprocessing for Urdu script may reduce these complexities and simplify the segmentation process. In character recognition system preprocessing intends to remove/reduce the noise, normalize image against present variations like skewness, slant, size etc. and minimize the storage requirement to increase processing speed. Preprocessing methods may be script oriented suitable techniques for Devnagari script may not work successfully for Arabic script due to different baseline, slants, dictates and modifiers etc. In case of Urdu and Arabic presence or absence of even a single dot can change the meaning of character and can be place in wrong character class. One of the major reasons of the segmentation dilemma in Urdu is due to unavailability of proper baseline because all character of a word doesn't join or aligned at one common line. An ideal preprocessing module normalizes all variation by applying noise removal, deskewing, normalization and image compression techniques.

2 Earlier Preprocessing Techniques

Due to its importance and the further results of feature extraction and recognition are strongly depends on preprocessing many researcher tried to introduce better preprocessing techniques and proposed various methodologies. Peters [1] presented a new method by using mathematical morphology for noise reduction. In his paper he described new morphological image cleaning algorithm (MIC) an residual image is recombined with processed residual images. Imran Razzak [2] Proposed a preprocessing method for both offline and online to obtain detail information for preprocessing by concentrating on smoothing, de-hooking, interpolation etc. Hussain [3] used smoothing technique and compute the displacement between two points using displacement from 4, 8, 16. Also [4] presented an approach for detecting orientation of Urdu documents images by different fonts and layouts using discriminative learning method. The method is based on classification of single connected components orientation present in a document image. Malik et al. [5] present an approach for Handwritten Urdu Word Spotting for this they used the method called connected components analysis and average filter. In [6] Saqib et al. presented Urdu document image layout analysis using multiresolution morphology based method, ridge detection and Gaussian filter bank smoothing methods. For smoothing and removal of salt and paper noise [7] used Median filters on binarized document images and later bounding box technique is used to remove noise. For Urdu document [8] used gray scale thresholding for removing noise and converted image into binary. Somaya et al. [9] performed normalization of strokes, slants, slope height and width of the letters as method proposed in [10] for baseline correction. Safwan [11] used median filter and minimum filter of special for preprocessing and removing jaggedness of the word and character contours. In [12, 13] present character recognition system which is robust to noise and scale invariant. In languages like Arabic, Urdu, Persian and Jawi the dots are very important by this consideration [14] proposed a preprocessing technique for Persian document which removes the noise similar to the dot and dicrats. Where smallest rectangle for each connected component are specified to decide threshold value. In [15] the threshold is determine by modified version of maximum entropy sum and entropic correlation method using 3×3 structuring element; Morphological closing flowed by morphological opening is performed to remove spurious segments [16]. Used explicit noise model in which these noise models are trainable automatically [17]. Suppress small island type noise which don't contribute shape information, block connectors and narrow channels is reduced using closing operation followed by opening operation. In [18] for noise reduction image is converted into monochrome image from which the noise element are removed. In [19] Cheung used Otus method for document binarization and [20] used word and subword method with two threshold values TN and TM. In this method the words and sub-words are extracted and their contours are classified by analyzing it. The contour's classes is found by setting threshold value TN and TM. In [21] used median filter of 3×3 and 5×5 size for Arabic document [22] mentioned linguistic topology rules and shows diacritics and dots are noise sensitive. In [23] smoothing followed by thresholding is used for reducing noise after thinning [24] used region growing technique for noise removal where linear regression method is used for removing the role lines printed on pages. Abuhaiba et al. [25] applied CBSA algorithm and noted that the CBSA is more better to tolerate the document image noise than thinning. In [26] Abdullah et al. for rejecting image noise the coefficient are used which comprise as large magnitude as possible. The Jawi script is also similar to Urdu script so the noise removal techniques used for Jawi may also work better for noise removal of Urdu characters. Mohammad Faidzul Nasrudin et al. [27] presented a fruitful review on Jawi handwritten character recognition in their work they review various proposed methods of [31–39] for preprocessing and normalization. Khairuddin et al. [28] used a Median filter is used then binary image is produce by threhsoding as given in [29]. And image compression is done by thinning algorithm of [30]. Mazani [31] used gamma and intensity correction as suggested by [32] and converted it grayscale by proposed method of [32, 33] and later used morphological filtering.

And finally thinning algorithms proposed in [34] are applied for compression of an image. In [35] a 3 × 3 structuring element is used and morphological closing and opening operation are performed. Mohd Sanusi [36] presented a reengineering method in which Gradient Orientation Histogram is used for slant and skew normalization, Nafiz et al. [37] categorized and presented a summary of all the character recognition methods systematically.

3 Preprocessing Objectives

Main objectives of preprocessing are shown in Fig. 1 which illustrates the purpose of preprocessing and it can be grouped under three major categories; noise reduction or Image enhancement, normalization, and compression to reduce the required amount of space and computation power [37]. The methods to achieve each objective are also shown under each objective in Fig. 1.

The first objective of preprocessing is noise reduction also called Image Enhancement. Many problems like broken character, overprinted characters, heavy noise, characters are normalized in preprocessing. Before feature extraction these imperfections are removed. Hundreds of available noise reduction techniques can be categorized in three major groups. [37–39]. In present paper only noise reduction using spatial filtering techniques are discussed, Normalization and compression techniques are beyond the scope of this paper.

To reduce such noise; various methods like filtering, Morphological operation and Noise modeling can be used. Image filtering methods are of two types spatial domain filter and frequency domain filters. The term spatial domain refers to the aggregate of pixels composing an image [40]. The main idea behind Spatial Domain Filtering is to convolve a mask with the image. Point processing, Filtering or Mask processing technique which are used for Smoothing and Sharpening of an

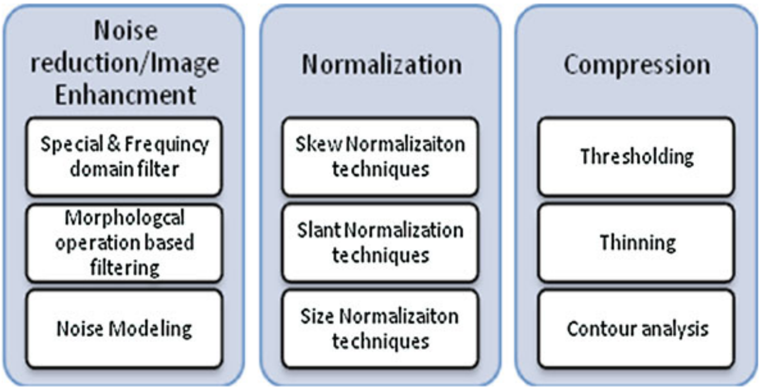


Fig. 1 Objectives of preprocessing

image. Low pass filter are used for smoothing; the smoothing is a noise reduction technique which is used to remove small details from an image. Median filter, Average filter, Max filter, Min Filter, etc. are used for smoothing. On the other hand sharpening is used to highlight the details of image like intensity details, variation details etc. One of the basic aims of Sharpening is to enhance the details of image which has been blurred during smoothing. First order derivatives and second order derivatives can be used for sharpening. Operator like Sobel, Prewitt masks are few of the best first order derivatives, similarly Laplacian is one of the best second order derivatives which is used as Laplacian of Gaussian i.e. LoG.

In present paper Spatial Lowpass filters like average filter, median filter, and Max filter are applied on Urdu document. Also histogram equalization and frequency domain Gaussian Lowpass filter are used to enhance the Urdu document image. *Averaging Filter* or Image averaging is used for smoothing the image, it is a special filtering technique where value of a particular pixel (x, y) is the average (arithmetic mean) of all the neighboring pixels. The smoothing is used to reduce the sharp transitions in the gray levels. The edges in image are blurred after applying average filter on it. Figure 2 shows a 3 × 3 mask where the constant multiplication for mask is equal to sum of the values of its coefficients.

$$g(x,y) = \frac{1}{9}\sum_{i=-1}^{1}\sum_{j=-1}^{1} f(x+i,\, y+j)$$

To avoid effect of blurring another kind of mask known as weighted average mask can be used. In this mask the pixel at central location will get the maximum weighted. And the weighted will decrease while moving away from the center of the mask like. Another filter is *Median Filter* the response of this filter is depends on ordering of the intensity values of pixels in the neighborhood of the pixel which is under consideration. For this filtering first take set of intensity values of all the neighborhood of (x, y). Then arrange all these intensity values in particular order. Based on these ordering, select median of these values and this will be the value of center pixel. The median filter is specially used for salt and paper noise. In *Maximum filter* each pixel of an image is replaced with the maximum value of neighborhood pixel. In this type of filter a mask like 3 × 3 is passed over the image and the central pixel of the image is assigned as the maximum value among all the neighborhood

Fig. 2 Different mask used in image smoothing (average filter)

$$W = \frac{1}{9} \times \begin{array}{|c|c|c|}\hline 1 & 1 & 1 \\ \hline 1 & 1 & 1 \\ \hline 1 & 1 & 1 \\ \hline\end{array} \qquad W = \frac{1}{16} \times \begin{array}{|c|c|c|}\hline 1 & 2 & 1 \\ \hline 2 & 4 & 2 \\ \hline 1 & 2 & 1 \\ \hline\end{array}$$

$$W = \frac{1}{4} \times \begin{array}{|c|c|c|}\hline 0 & 1 & 0 \\ \hline 1 & 0 & 1 \\ \hline 0 & 1 & 0 \\ \hline\end{array} \qquad W = \frac{1}{4} \times \begin{array}{|c|c|c|}\hline 0 & 1 & 0 \\ \hline 1 & 0 & 1 \\ \hline 0 & 1 & 0 \\ \hline\end{array}$$

the pixels. *Frequency domain Gaussian Lowpass filter:* When an image is to be processed in the frequency domain, the digitized image will first multiply the by $(-1)^{x+y}$ to center the transform. After multiplication transform the resultant image into frequency domain by using Discrete Fourier Transform DFT. Then perform the process on it using filter function in frequency domain. It can be defined by

$H(u,v) = e^{-D^2(u,v)/2D_0^2}$ Where D_0 is cutoff of frequency

Histogram equalization: Histogram equalization automatically determines a transformation function that seeks to produce an output image that has a uniform histogram. When automatic enhancement is desired then histogram equalization is a good approach. It is a technique for adjusting image intensities to enhance contrast. In histogram equalization, the input pixel intensity, r is transformed to new intensity value, s by transformation function T.

4 Experimental Work

In proposed system various filtering techniques are tested to find the best one for Urdu word document. Various types of noisy Urdu images are used to verify the filtering techniques. The noisy documents are categorized into two groups. First types of documents are very dark and noisy document which are difficult to read and understand even by human eyes. These document images are produced due to scanning devices of small dynamic range, very dark page background and poor document pages etc. which are shown in Fig. 3a. For such document image Algorithm I is proposed. And second category of document consists of those documents which contain noise but they are not like extremely dark and difficult to understand. As shown in Fig. 4a. For such Document image Algorithm II is proposed. These algorithms are given below.

Algorithm I–Steps to remove noise from Type 1 Document image:

1. Convert the document image into gray scale.
2. Use Histogram equalization
3. Apply 3 × 3 Max filter on resultant image of step (2)
4. Apply thresholding on result of step (3)
5. Apply Frequency domain Gaussian Lowpass filter on result of (4)
6. Noise free image.

Figure 3 illustrate the above algorithm for noise removal from dark and noisy image in which (a) shows original document image with which is dark. (b) is histogram equalized image of original image. In (c) and (d) a 3 × 3 Median and Max filters of (b) are shown respectively. It can be observed that the 3 × 3 Max filter is giving better results than Median filter in case of dark document image. (e) is showing thresholded binary image of d. And finally Frequency domain Gaussian Lowpass filter is applied on (e) which removes the high intensity pixles and blurred the image which produces a noise free document image as sown in (f).

Fig. 3 **a** Original image. **b** Histogram equalized image. **c** 3 × 3 median filter of image **b**. **d** 3 × 3 max filter of image **b**. **e** Thresholded image of **d**. **f** Frequency domain Gaussian filter of image **e**

Algorithm II–Steps to remove noise from Type 2 Document image:

1. Convert the document image into gray scale.
2. Apply 3 × 3 Max filter on gray scale image
3. Apply thresholding on result of step (2)
   a. If image consists high intensity foreground pixels then Apply Frequency domain Gaussian Lowpass filter on result of (3).
   b. Else go to step 4
4. Obtained Noise free image.

Fig. 4 *Left* hand column shows noisy images and *right* hand side column are results of algorithm II

Figure 4 illustrate the results of above algorithm in case of the noisy image without very dark background. In Fig. 4 Left column shows the noisy image and right side column shows the resultant image after applying Algorithm II on noisy images.

5 Conclusion

In present paper various preprocessing techniques for Arabic, Persian, Urdu and Jawi script were studied. Also an attempt was made to perform image enhancement of Urdu document image. Couples of algorithms are proposed for Noisy document image and very dark noisy document image. Spatial and frequency domain filters are used for noise removal. The Max and Median filters of 3×3 neighborhood are used for filtering. Also histogram equalization technique is applied on dark images for normalization purpose. Frequency domain Gaussian Lowpass filter is used for smoothing of the resultant image. The future work will continue toward slant, skew and size normalization.

Acknowledgments This work is sponsored by a G.H. Raisoni Doctoral fellowship, North Maharashtra University, Jalgoan. The author would like to acknowledge for their financial support.

References

1. Peters RA (1995) A new algorithm for image noise reduction using mathematical morphology. IEEE Trans Image Process 4:554–568
2. Imran Razzak M, Afaq Hussain S, Muhammad S, Shafi Khan Z (2009) Combining offline and online preprocessing for online Urdu character recognition. In: Proceedings of the international multiconference of engineers and computer scientists, vol 1, Hong Kong, 18–20 March 2009, IMECS 2009
3. Hussain M et al (2005) Urdu character recognition using spatial temporal neural network. In: Proceedings of 9th international multitopic conference, IEEE INMIC 2005, 24–25 Dec 2005, pp 1–5
4. Sheikh Faisal R, Syed Saqib B, Faisal S, Thomas MB (2009) A discriminative learning approach for orientation detection of Urdu document images. In: Proceedings of multitopic conference, IEEE 13th international, INMIC 2009, 14–15 Dec 2009, pp 1–5
5. Malik Waqas S, Nicola N, Lei He C, Ching YS (2010) A novel handwritten Urdu word spotting based on connected components analysis. In: Proceeding of 2010 international conference on pattern recognition, IEEE computer society 2010
6. Syed Saqib B, Faisal S, Thomas MB (2011) High performance layout analysis of Arabic and Urdu document images. In: Proceedings of 2011 international conference on document analysis and recognition, pp 1275–1279
7. Malik Waqas S, Nicola N, Chun LH, Ching YS (2010) Holistic Urdu handwritten word recognition using support vector machine. In: Proceedings of 12th international conference on pattern recognition, 23–26 August, ISBN: 978-0-7695-4109-9
8. Shuwair S, Abdul W (2010) Optical character recognition system for Urdu, Information and emerging technologies (ICIET). In: Proceedings of 2010 international conference 9 Nov 2010
9. Alama' adeed S et al. (2002) Recognition of offline handwritten Arabic word using hidden Markov model approach, ICPR 02. In: Proceedings of the 16th international conference on pattern recognition (ICPR 02), vol 3, ISBN:0-7695-1695-X, p 30481
10. Pechwitz M, M̈argner V (2002) Baseline estimation for Arabic handwritten words. In: Proceedings of 8th international workshop on frontiers in handwriting recognition, IWFHR 2002, August 2002, Niagara-on-the-Lake, Canada 2002
11. Safwan W, Zhixin S,Venu G (2009) Segmentation of Arabic handwriting based on both contour and skeleton segmentation. In: 10th international conference on document analysis and recognition, pp 793–797
12. Pal U, Anirban S (2003) In: Proceedings of 7th international conference on document analysis and recognition, pp 1183–1187
13. Atique Ur Rehman M (2010) A new scale invariant optimized chain code for Nastaliq character representation. In: 2nd international conference on computer modeling and simulation, pp 400–403
14. Hassan Shirali-Shahreza M, Shirali-Shahreza S (2008) Removing noises similar to dots from persian scanned documents. In: ISECS international colloquium on computing, communication, control, and management, pp 313–317
15. Vaseghi B, Alirezaee S, Ahmadi M, Amirfattahi R (2008) Off-line Farsi/Arabic handwritten word recognition using vector quantization and hidden Markov model. In: Proceedings of multitopic conference, INMIC 2008. IEEE International, 23–24 Dec 2008, pp 575–578
16. Al-Badr B, Robert MH (1995) Segmentation-free word recognition with application to Arabic. In: Proceedings of the 3rd international conference on document analysis and recognition, ICDAR '95, pp 355–359

17. Deya M, Adnan A, Robert S (1997) Segmentation of Arabic cursive script. In: Proceedings of the 4th international conference on document analysis and recognition pages, ICDAR '97, pp 625–628
18. Muhammad S, Syed Nazim N, Abdulaziz A-K (2003) Offline Arabic text recognition system. In: Proceedings of the 2003 international conference on geometric modeling and graphics GMAG-03
19. Cheung A, Bennamoun M, Bergmann NW (2001) A recognition-based Arabic optical character segmentation. Pattern recognition, vol 34, pp.215–233
20. Kavianifar M, Adnan A (1999) Preprocessing and structural feature extraction for a multi-fonts Arabic/Persian OCR. IJDR 1999, pp 213–216
21. Mahmoud AL-Shatnawi A, AL-Salaimeh S, AL-Zawaideh FH, Khairuddin O (2011) Offline Arabic text recognition——an overview. World Comput Sci Inform Technol J (WCSIT) ISSN: 2221-0741 1(5): 184–192
22. Ahmed ME, Mohamed AI (2001) A graph-based segmentation and feature extraction framework for Arabic text recognition. In: Proceedings of ICDAR 2001, pp 622–626
23. Ziad O, Lebanon B (2009) Automatic processing of Arabic text. In: Proceedings of the 6th international conference on Innovations in information technology, IIT'09, p 6–10
24. Zhixin S, Srirangaraj S, Venu G (2011) Image enhancement for degraded binary document images. In: Proceedings of international conference on document analysis and recognition, pp 895–899
25. Abuhaiba ISI, Mahmoud SA, Green RJ (1994) Recognition of handwritten cursive Arabic characters. IEEE Transactions Pattern Anal Mach Intell 16(6): 644–672
26. Al-Shoshan AI (2006) Arabic OCR based on image invariants. In: Proceedings of the geometric modeling and imaging trends (GMAI'06)—July 2006 New
27. Mohammad FN, Khairuddin O, Mohamad SZ, Liong CY (2010) Handwritten cursive Jawi character recognition: a survey. In: Proceedings of 5th international conference on computer graphics, imaging and visualization, pp 247–256
28. Khairuddin O (2000) Jawi handwritten text recognition using multi-level classifier (in Malay), PhD Thesis, Universiti Putra Malaysia
29. Sharaf El-Deen S, Horaini M, Zainodin J, Khairuddin O (1993) Skeletonization, Laporan Teknik Jabatan Sains Komputer. Fakulti Sains Matematik dan Komputer. Universiti Kebangsaan Malaysia, Bangi
30. Naccashe NJ, Shinghal R (1984) SPTA: a proposed algorithm for thinning binary patterns. IEEE Trans Syst Man Cybernatics, SMC-14(3), May/June, pp 409–418
31. Mazani M (2002) In: Jawi handwritten text recognition using recurrent Bama neural networks (in Malay), PhD thesis, Universiti Kebangsaan Malaysia
32. Parker JR (1994) Practical computer vision. Wiley, Singapore
33. Philips D (1994) Image processing: analyzing and enhancing digital images. R&D Publications Inc, Kansas
34. Zhang TY, Suen CY (1984) A fast algorithm for thinning digital pattern. Comm ACM 7(3):236–239
35. Chen M-Y, Kundu A, Zhou J (1994) Off-line handwritten word recognition using a hidden Markov model type Stochastic network. IEEE Trans Patter Analy Mach Intell 16(5):481–496
36. Mohd Sanusi A (2003) Reengineering of slant and slope orientation skew histogram for Merong Mahawangsa Manuscript (in Malay), MIT Thesis, Fakulti Teknologi dan Sains Maklumat, Universiti Kebangsaan Malaysia, Bangi
37. Nafiz A, Yarman-Vural FT (2001) An overview of character recognition focused on off-line handwriting. IEEE Trans On Patter Analysis and Machine Intelligence 31(2): 216–233
38. Serra J (1994) Morphological filtering: an overview. Signal Process 38(1):3–11
39. Sonka M, Hlavac V, Boyle R (1999) Image processing, analysis and machine vision, 2nd edn. Brooks/Cole, CA
40. Gonzalez RC, Woods RE (2004) Digital image processing, 2nd edn Pearson Education

Medical Image Spatial Fusion Watermarking System

P. Viswanathan and P. VenkataKrishna

Abstract The watermarking based on wavelet fusion gains sharp edge and discontinuity in watermarked image due to embedding in smooth region. To solve this problem spatial fusion watermarking method is proposed in this paper. The image is decomposed into four levels by predicting the range of intensity. To improve the capacity of embedding, higher counted region is chosen for embedding which indirectly chooses sharp region. The Text or ROI data is watermarked in the chosen region by differentiating the pixel with one by the constraint of even or odd between the cover image and text or ROI. The decomposed images are composed by averaging the pixels of the regions greater than one. After extraction, the watermarked medical image is reconstructed to original medical image by reversible property. This system is evaluated with various metrics using standard medical images which show good quality and high imperceptibility with embedding capacity.

Keywords Fusion · Watermarking · Decomposition · Extraction · Composition and reversible

P. Viswanathan (✉)
Infromation Technology, VIT University, Vellore, India
e-mail: pviswanathan@vit.ac.in

P. VenkataKrishna
Computer Science Engineering, VIT University, Vellore, India
e-mail: pvenkatakrishna@vit.ac.in

Mohan S. and S. S. Kumar (eds.), *Proceedings of the Fourth International Conference on Signal and Image Processing 2012 (ICSIP 2012)*, Lecture Notes in Electrical Engineering 222,
DOI: 10.1007/978-81-322-1000-9_43,

1 Introduction

In today's world, for patients have large amount of medical transcription maintained online requiring more time to upload, perfect authentication and reconstruction of image is needed to avoid wrong diagnose. The patient or hacker may also grab the unauthorised information leading to medical insurance forgery affecting the privacy and confidentiality of the patient [1–3]. The multimodal medical image is the process of handling medical image and text information of the patients. The text or ROI of the patients is hidden as a watermark in the medical image to improve security level and reduce the amount of maintenance of medical data.

The Fusion is a technology of combining more information in a single scene to provide comprehensive and reliable description or interpretation, for segmentation, feature extraction and target recognition. The requirements of image fusion algorithm are

1. Avoid elimination of salient information in the image.
2. Avoid introduction of artifacts or inconsistencies in image which leads to distract.
3. Reliable and robust against noise.

Pyramid [4] and wavelet [5, 6] are the most widely used image decomposition algorithms including Gradient pyramid and Ratio pyramid. During decomposition, the image is divided into several parts that represent different meaning. The traditional pyramid based decomposition calculates the approximation of an image and obtains the prediction error using source image. The decomposition using wavelet is by low pass band and high pass band. During composition stage, the approximation and detailed coefficient is treated using fusion rules. Detail coefficient having large absolute values denote the salient features in the image such as edges, lines and region boundaries. Approximate coefficient represents the mean intensity or some coarse texture information of the image [7–11].

The detail image composition is performed by selection of large detail image values called as selective combination based on the calculation of energy using the decision map of the detail image fusion rules. It gains sharp edge and discontinuity in fused image. Hence the window is used to calculate energy [11]. It improves the continuity of the image and reduces the influence of salt and pepper noise. However, there is a need of consistency which can be solved by our proposed approach.

The Label Ranking means the ranking of information mined based on correlative visual information [12, 13]. Discrete wavelet transform which satisfies the Orthonormal property divide the image into high frequency and low frequency components and then hide the watermark in the component. It gives efficient reconstruction of the image after extraction of watermarking.

Table 1 Key notation of image

MI	Medical image
M_R	One dimensional row vector of medical image
L_i	Limit estimation of index i
D_d	Decomposed image of label d = P, Q, R, S
D_d^m	Maximum count decomposed region
P, Q, R, S	Representation of average horizontal, vertical and diagonal information M
M, N	Size of image
I_d	Data fused
M_f	Fusion medical image

1.1 Contribution and Scope

1. Present a fusion watermarking scheme designed by the context of range of intensity level.
2. Develop a method to embed the information with high imperceptibility.
3. Develop a method to extract the information without loss of data.
4. The performance of the proposed medical image fusion watermarking is evaluated in terms of quality, imperceptibility.

The notation used to evaluate the algorithm is shown in Table 1 contains the keyword denote the various aspects of medical image fusion.

This paper is organised as Multimodal medical image fusion watermarking in Sect. 2 and the extraction of the image or information with reversible property is in Sect. 3. The performance evaluation is in Sect. 4 and conclusion in Sect. 5 respectively.

2 Multimodal Medical Image Spatial Fusion Watermarking

The Multimodal medical image spatial fusion watermarking is the embedding of text or region of interest data in the medical image based on fusion principle. The overview of the proposed approach for spatial fusion watermarking of text or ROI with medical image is shown in Fig. 1. The Multimodal medical image spatial fusion watermarking is organized as the limit estimation and formulation in Sect. 2.1, decomposition in Sect. 2.2 and the fusion of information with composition is in Sect. 2.3 (Fig. 2).

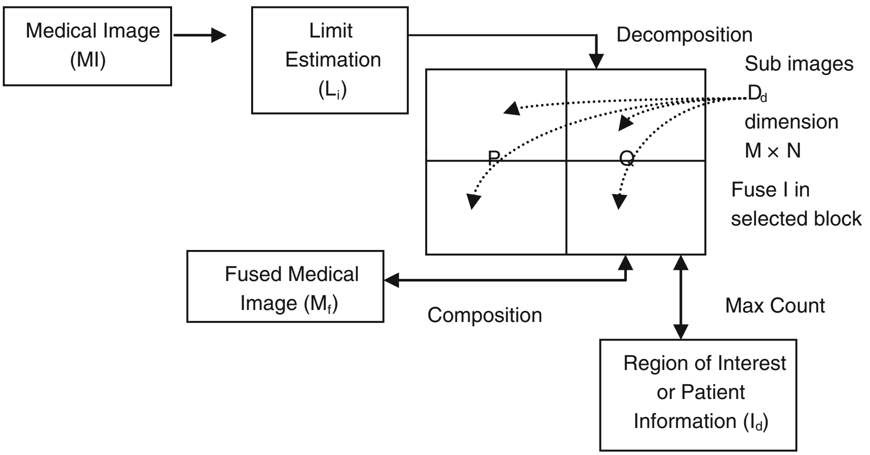


Fig. 1 Overview of proposed medical image spatial fusion watermarking

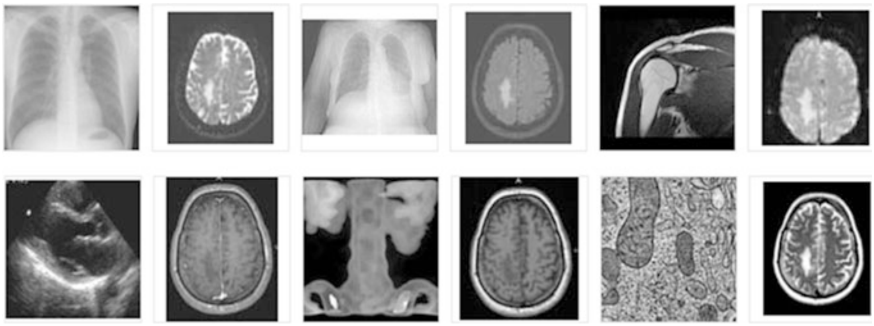


Fig. 2 Original medical image

2.1 Formulation of Limit

The medical image is analysed and decomposed by predicting the level of intensity. The labels of image are denoted by $L_i \in M$ with 0 or 1. The label of the image is evaluated by the range of intensity level in the image. The medical image is represented in a one dimensional row vector sorted in ascending order denoted by $M_R \in M$. The limits L_i are estimated by defining the location of the image vector M_R using Eq. (1)

$$\mathop{\forall}_{i=1}^{3} L_i = MI_R(l \times i)\, where\, l = \frac{N}{4} - 1 \tag{1}$$

$N = r \times c$ size of the image, r no of rows, c no of columns of the image and L_i represent the label and the limit of corresponding part for the image. It provides the effective representation of the medical image. The algorithm (1) is used for estimating the limits represent different regions of the medical image. The accurate prediction of limits varies depend upon the image hence the system provides automotive evaluation for level of decomposition.

Algorithm 1: Limit estimation
Input: Medical image
Output: Limit values
Steps:
1. Estimate the size of the image//N = size (MI)
2. Define the medical image as a single row vector//M_R = M(x, y)
3. Sort the vector M_R in ascending order
4. Divide the estimated size of the image with the level of decomposition//d = N/4
5. Estimate the limits by pointing the location of the medical image
// L1 = R(d) // L2 = R(d * 2) // L3 = R(d * 3)

2.2 Decomposition of Image

The image is decomposed by the estimated limit from algorithm 1. The limits act as a splitter to decompose the medical image into different region D_d by using Eq. (2).

$$D_d(x,y) = \sum_{x=0}^{N}\sum_{y=0}^{N} MI(x,y) < L_i \quad count_i + + \tag{2}$$

where,

$$i = 1\, then\, D_S(x,\, y)$$

$$i = 2\, then\, D_R(x,\, y)$$

$$i = 3\, then\, D_Q(x,\, y)$$

$$else\, D_p(x,y)$$

where d represents the label by means of P, Q, R and S, which determines the different region of image. The count of each region is evaluated to determine the maximum counted level region for embedding the text information or ROI, I_d. The algorithm (2) decompose the medical image into various level of the image shown in Fig. 3. It shows the different parts of image with varying intensity level provides accurate representation of medical image gives accurate diagnose of medical data.

Algorithm 2: Decomposition of image
Input: Medical image and Limit values
Output: Sub images P, Q, R, S

(continued)

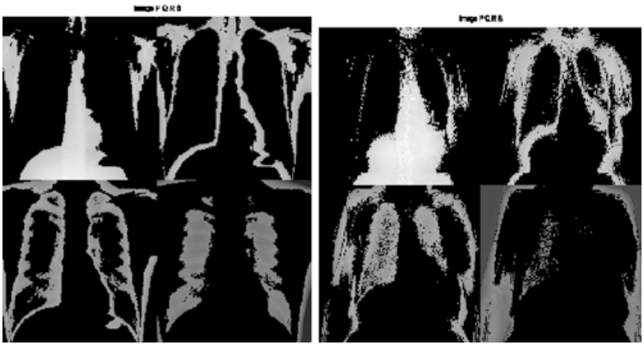


Fig. 3 Decomposed images

(continued)

Algorithm 2: Decomposition of image
Steps:
1. The estimated limit is compared with the medical image
2. Check the intensity less than the limit
3. The level of image is estimated and counted the no of pixel
4. First range level of the image D_S//MI < L_1//count(4) + 1
5. Second range level of the image D_R//MI < L_2//count(3) + 1
6. Third range level of the image D_Q//MI < L_3//count(2) + 1 else
7. Fourth range level of the image D_P//count(1) = count(1) + 1

2.3 Watermarking and Composition

Watermarking is the process of hiding information in Image. The information or ROI of patient I_d is hidden in the medical image by evaluating the maximum counted region D_d^m from the decomposed region, estimated using algorithm (2). The information or data I_d is merged in particular region D_d^m by the differentiating the pixel by one in the intensity level between the text information or ROI and the medical image even or odd by using Eq. (3).

$$D_d = D_d^m - 1 \; if \; D_d^m \neq 0 \begin{cases} I_d = 0 \&\& D_d^m \, \%2 \neq 0 \\ I_d \neq 0 \&\& D_d^m \, \%2 = 0 \end{cases} \tag{3}$$

A decomposed region are composed into a medical image with the watermarked data is shown in Fig. 4. The decomposed regions are added into single medical image using the Eq. (4). It shows that the image obtained have no variation between the original medical image.

$$F(x,y) = \sum_{x=0}^{N} \sum_{y=0}^{M} D_d(x,y) > 1 \tag{4}$$

Algorithm 3: Watermarking and Composition of image
Input: Medical image, Limit values and ROI or information
Output: Watermarked image
Steps:
1. Estimate the limit using algorithm 1
2. Decompose the image into region and max counted region D_d^m is estimated using algorithm 2
3. Check intensity level of the region not equal to zero
4. Merge the information or ROI, I_d in the medical image
5. If $I_d = 0$ and D_d^m is odd differentiate the D_d^m with -1
6. If $I_d! = 0$ and D_d^m is even differentiate the D_d^m with -1
7. $D_P \Vert D_Q \Vert D_R \Vert D_S = D_d^m$
8. Mean of the region is checked greater than one
9. The decomposed image is composed into a medical image
10. The Watermarked medical image, $M_F = D_P + D_Q + D_R + D_S > 1$

The algorithm (3) for the fusion of medical image with data satisfies the requirement of the fusion rules by means of composition and decomposition without any loss in the medical image.

3 Extraction of Information or Region of Interest from Image with Reversible

The information or data from the watermarked image is extracted to provide authentication of the medical image. The fusion watermarked image obtained from algorithm (3) is again evaluated by algorithm (1) and (2) to estimate the limit and to decompose the image into regions. The max counted decomposed region D_d^m is evaluated to extract the information or ROI from fusion image by using Eq. (5)

$$I_d = \begin{cases} 0 & D_d^m > 1\&\&\%2 = 0 \\ 1 & D_d^m > 1\&\&\%2 \neq 0 \end{cases} \quad (5)$$

The algorithm (4) extracts the merged data from the watermarked image. The obtained copyright image and text is shown in Fig. 5 shows efficient extraction of information I_d from the fusion watermarked image without any loss.

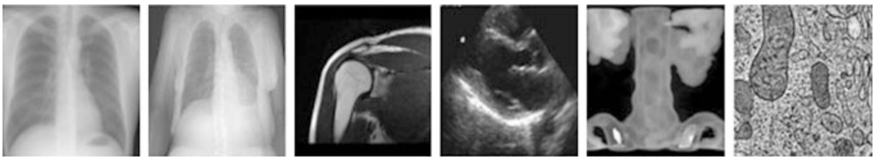


Fig. 4 Watermarked image

Algorithm 4: Extraction

Input: Watermarked image
Output: ROI or Text
Steps:
1. Estimate the limit using algorithm 1
2. Decompose the image into region and max counted region D_d^m is estimated using algorithm 2
3. Check D_d^m greater than one
4. If D_d^m is even than initialize the fission value 0
5. If D_d^m is odd than initialize the fission value 1
6. Convert the Fission value to decimal determines the data merged

---

The reversible operation is performed during the extraction process using the reversible property obtained during fusion provide original medical source image using Eq. (6)

$$D_d = D_d^m + 1 \; if \; D_d^m \neq 0 \tag{6}$$

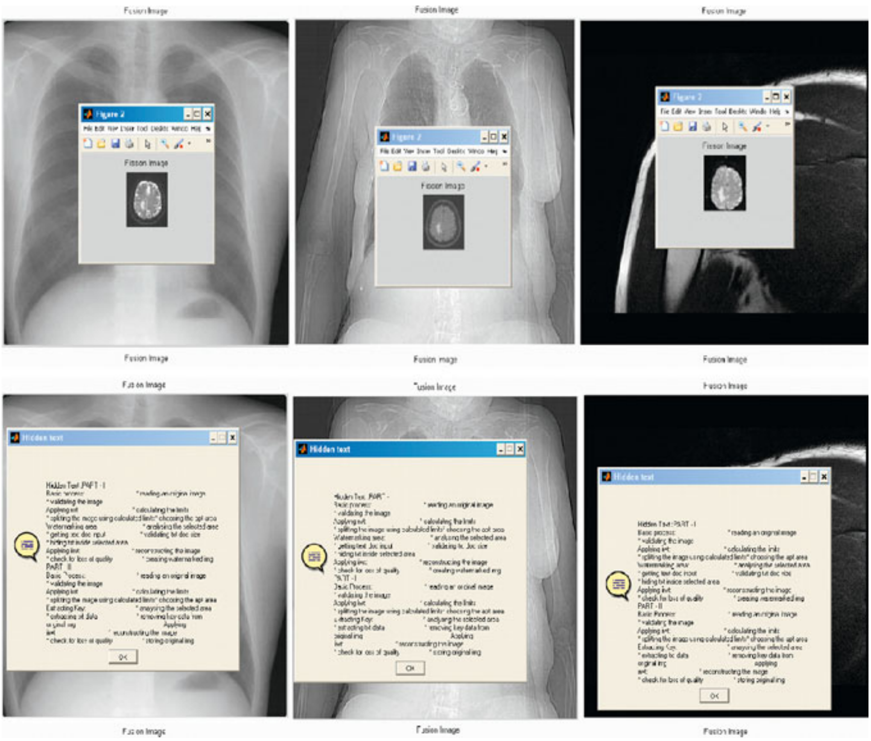


Fig. 5 Extraction of the image and text from the medical image

4 Performance Evaluation

The proposed approach is evaluated by the medical image benchmarks of dataset by DICOM [18] and IEEE visualization context [19] shown in Fig. 2. The DICOM images are taken as cover image and IEEE image as ROI and text for hiding. The standard metrics like peak signals noise ratio, Root mean squared error, signal to noise ratio etc. [14–17] to determine the quality of the fusion watermarked image. The computed Root mean square error (RMSE) and percentage fitness error (PFE) shown in Fig. 6a is very low hence there is no difference. The PFE for CT, CR, XRAY and SCAN provides less error rate than MR and US images resulted very less deviation between the original medical image and fusion image compare to MR and US fusion image. The Mean square error of the extracted image is zero determine that no variation between the Fused image and fission image means, it resembles or identical with the original copyright image.

The Signal to Noise Ratio (SNR) and Peak signal noise ratio (PSNR) is estimated between source medical image with fusion watermarked image with image

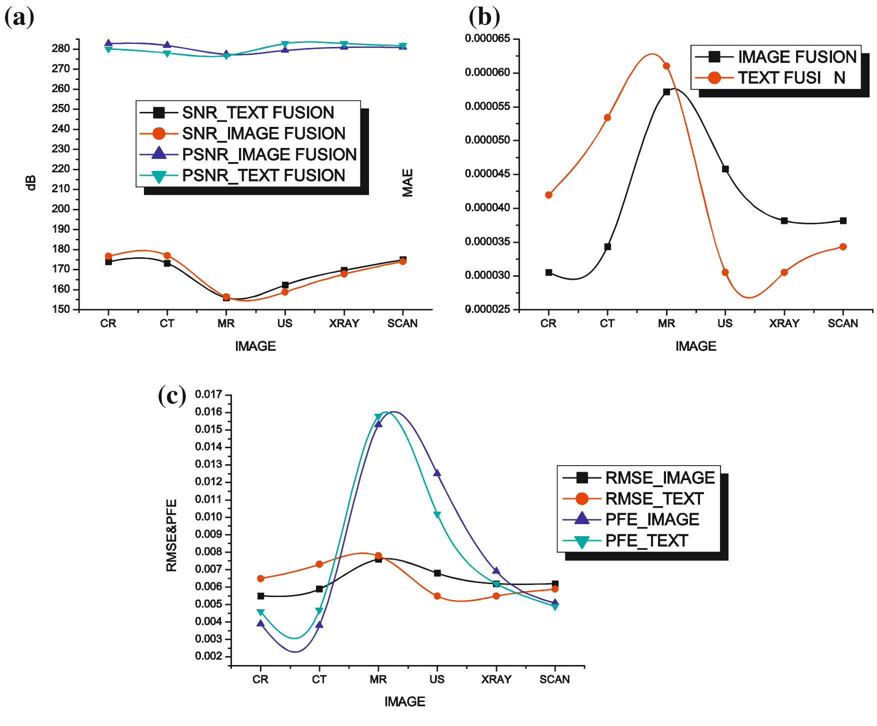


Fig. 6 **a** Peak signal noise ratio and signal noise ratio between image and text fused image. **b** Mean absolute error between image and text fused image. **c** Root mean square error (RMSE) and percentage fitness error (PFI) between original medical image and fusion image with text and image

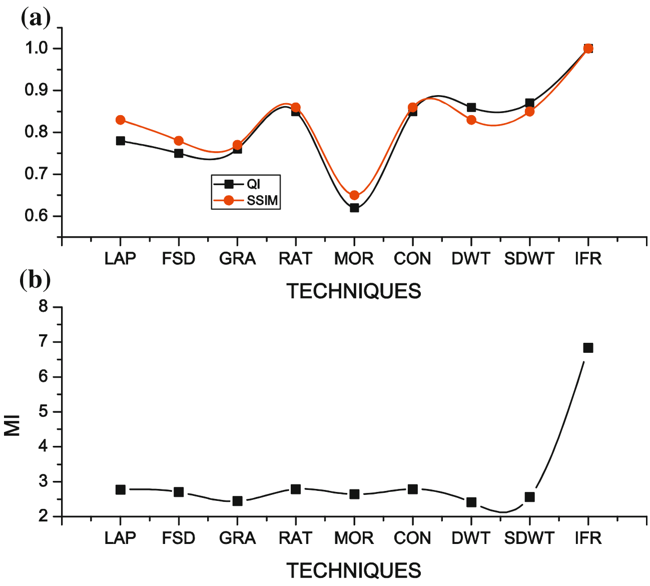


Fig. 7 **a** Universal quality index (QI) and structure similarity (SSIM) between original medical image and fusion image. **b** Mutual information between original medical image and fusion image

and text shown is in Fig. 6b resulted large value means that the sensitivity of medical image is maintained. The mean absolute error (MAE) evaluation in Fig. 6c has shown very less error which determines the close prediction between the medical image and fusion image.

The proposed method performed comparably better than the algorithm like Laplace (LAP), Filter Subtract decimate (FSD), Gradient (GRA), Ratio (RAT), Morphological (MOR), Contrast (CON), Discrete Wavelet Transform (DWT) Shift invariant Discrete Wavelet Transform (SDWT) are evaluated with the metrics of mutual information (MI), Universal quality index (QI) and Structure similarity (SSIM) is shown in Fig. 7a and b).

The MI is large when compared to other techniques tells that image information is not affected more than the previous approaches. The QI, of most of the images reached one means the quality of medical image is maintained and SSIM also high determine the maximum structure similarity between the fusion and original medical image. From the performance evaluation there is no major effect of fusion in medical image and extracted image resembles the original image. Hence the algorithm maintained the quality of the image with efficient fusion watermarking.

5 Conclusion

The Medical Image Spatial Fusion Watermarking in region of range based contest improves the fusion of multi scale image decomposition and composition. The proposed fusion watermarking process provides no loss of data because it provides higher PSNR than the wavelet domain and decomposition is localized in spatial and frequency domain like wavelet. The drawback is more level of decomposition cannot be performed like the wavelet but the different level of the image provides the various aspects of the medical image used for efficient diagnose. Experimentally we found that the fusion of ROI or information in the image is highly imperceptible in the medical image can be used for the security and authentication of medical image in order to avoid forgery of medical insurance.

References

1. Viswanathan P, Venkata Krishna P (2009) Text fusion watermarking in medical image with semi-reversible for secure transfer and authentication. In: International conference on advances in recent technologies in communication and computing. ARTCom '09, pp 585–589, 27–28 Oct 2009
2. Viswanathan P, VenkataKrishna P (2011) Fusion of cryptographic watermarking medical image system with reversible property. ICTACT Int J Image Video Process 2(1):258–263
3. Gutierrez CN, Kakani G, Verma RC, Wang T (2010) Digital watermarking of medical images for mobile devices. In: IEEE international conference on sensor networks, ubiquitous, and trustworthy computing (SUTC), pp 421–425, 7–9 June 2010
4. Tu ZW, Chen XR, Yuille AL, Zhu SC (2005) Image parsing: unifying segmentation, detection, and recognition. Int J Comput Vis 63(2):113–140
5. Burt PJ, Adelson EH (1983) The Laplacian pyramid as a compact image code. IEEE Trans Commun 31(4):532–540
6. Pajares G, de la Cruz JM (2004) A wavelet-based image fusion tutorial. Pattern Recog 37(9):1855–1872
7. Ray LA, Adhami RR (2006) Dual tree discrete wavelet transform with application to image fusion. In: Southeastern symposium on system theory, pp 430–433
8. Xydeas CS, Petrovic V (2004) Gradient-based multiresolution image fusion. IEEE Trans Image Process 13(2):228–237
9. Toet A (1989) Image fusion by a ratio of low-pass pyramid. Pattern Recogn Lett 9:245–253
10. Toet A, van Ruyven LJ, Valeton JM (1989) Merging thermal and visual images by a contrast pyramid. Opt Eng 28(7):789–792
11. Simoncelli EP, Freeman WT, Adelson EH, Heeger DJ (1992) Shiftable multi-scale transforms. IEEE Trans Inf Theor 38(2):587–607
12. Qu GH, Zhang DL, Yan PF (2002) Information measure for performance of image fusion. IEEE Electron Lett 38(7):313–315
13. Lifeng Y, Donglin Z, Weidong W, Shanglian B (2001) Multi-modality medical image fusion based on wavelet pyramid and evaluation. Syst Eng Electron 12:42–48
14. Wang Q, Shen Y (2006) Performance assessment of image fusion. In: Advances in image and video technology, Springer, Berlin, pp 373–382
15. Wang Z, Bovik AC, Sheikh HR, Simoncelli EP (2004) Image quality assessment: from error measurement to structural similarity. IEEE Trans Image Process 13(4):600–612

16. Wang Q, Shen Y (2006) Performance assessment of image fusion. In: Advances in image and video technology, Springer, Berlin, pp 373–382
17. Xydeas CS, Petrovic V (2000) Objective image fusion performance measure. Electron Lett 36(4):308–309
18. DICOM image data download http://barre.nom.fr/medical/samples/
19. IEEE visualization context data download http://viscontest.sdsc.edu/2010/data.html

Classification of Medical Imaging Modalities Based on Visual and Signal Features

Amir Rajaei, Elham Dallalzadeh and Lalitha Rangarajan

Abstract In this paper, we present an approach to classify medical imaging modalities. Medical images are preprocessed in order to remove noises and enhance their content. The features based on texture, appearance and signal are extracted. The extracted features are concatenated to each other and considered for classification. KNN and SVM classifiers are applied to classify medical imaging modalities. The proposed approach is conducted on IMageCLEF2010 dataset. We achieve classification accuracy 95.39 % that presents the efficiency of our proposed approach.

Keywords Medical imaging modalities classification · Feature extraction · Texture feature · Appearance feature · Signal feature · K-nearest neighbor · Support vector machine

1 Introduction

Computer Aided Diagnosis (CAD) plays an important role in the field of diagnosis and clinical care. To have precise diagnosing of specific diseases, medical imaging modalities are widely used. In CAD, the basic objective is to obtain the images

A. Rajaei (✉) · E. Dallalzadeh · L. Rangarajan
Department of Studies in Computer Science, Manasagangothri, University of Mysore, Mysore 570006, India
e-mail: rajaei80amir@gmail.com

E. Dallalzadeh
e-mail: elhamdallalzadeh@gmail.com

L. Rangarajan
e-mail: lali85arun@yahoo.co.in

Mohan S. and S. S. Kumar (eds.), *Proceedings of the Fourth International Conference on Signal and Image Processing 2012 (ICSIP 2012)*, Lecture Notes in Electrical Engineering 222,
DOI: 10.1007/978-81-322-1000-9_44,

related to relevant past cases. Different medical imaging modalities such as X-ray, CT, MR, PET, Ultrasound, and Micro are used in CAD [1].

On the other hand, huge amounts of medical imaging modalities are stored in the datasets. Automatic classifications of medical images are vastly needed. In this direction, efficient and reliable automatic classification methods are necessary to categorize the large medical images with less human intervention [2].

In literature, several attempts are reported on classification of medical imaging modalities. Florea et al. [3] proposed an approach to classify medical imaging modalities based on textual annotation interpretation. The textual annotations are interpreted using a set of 96 production rules defined by expert radiologists. Classification of medical images based on their modalities and body organs are proposed by Malik and Zermic [4]. They have introduced an algorithm based on a look up table and energy information obtained from Wavelet transform in order to classify medical images based on modalities and subsequently body organs. Look up table is used to identify the image modalities. Different body organ are then classified by extracting and representing the Biorthogonal and Reversible Biorthogonal wavelet transform.

Florea et al. [5] proposed to capture the spatial description of features inside the images. The medical images are divided into 26 equal blocks. Haralicks coefficients are then extracted from the Gray-Level Co-occurrence Matrix (GLCM) to characterize medical image texture. Moreover, the medical image texture is described by box counted fractal dimension and Gabor wavelet. In addition, they estimated first, second, third, and fourth moments of the obtained features. Finally, KNN classifier is applied for classification of medical imaging modalities.

Han and Chen [6] proposed an approach to classify medical imaging modalities by extracting visual features as well as textual features. Visual features are extracted in terms of global and local features. Histogram descriptor of edges, color intensity, and block-based variations are extracted and represented as global features. They used SIFT histogram as local features. The binary histograms of some predefined vocabulary words of image caption are extracted and represented as textual features. All the extracted features are then concatenated to each other. They applied SVM classifier to classify medical imaging modalities.

Kalpathy and Hersh [7] proposed an approach to classify different modalities of medical images based on concatenation of texture and histogram features. First, each image is resized into 256×256 and then divided into five non-overlapping blocks. The Gray Level Co-occurrence Matrix (GLCM) of each block is obtained. From the obtained matrix, a set of texture features such as contrast, correlation, energy, homogeneity, and entropy are extracted. The histogram of each block is also extracted. The gray scale histograms are concatenated with texture features. Neural Network classifier is then applied to classify the medical imaging modalities. Pauly et al. [8] proposed an approach to extract the visual features such as histograms, scale-space, Log-Gabor, and phase congruency of medical images as feature extraction. SVM and Random Forest classifiers are then applied to classify medical images.

From the literature survey, it is observed that some of the classification approaches are based on textual features. Textual features are obtained by the experts' interpretation. Physicians and radiologists may have different observation about specific diseases. In addition, the resolutions of medical images are very low. Therefore, color features are not robust and also they are with strong noise. Thus, it is necessary to represent medical images by extracting efficient feature without expert intervention.

In this paper, we propose to extract the visual and signal features. Texture and appearance feature are extracted as visual features along with extracted signal features. Due to the different source of radiation of the devices that capture medical images are having particular visual and signal characteristics. K Nearest Neighbor (KNN) and Support Vector Machine (SVM) classifiers are applied to classify medical imaging modalities.

The rest of the paper consists of following sections. Section 2 presents our proposed approach for classification of medical imaging modalities. The details of dataset conducted to evaluate the performance of the proposed approach along with experimental results are discussed in Sect. 3. The paper is concluded in Sect. 4.

2 Proposed Approach

In this section, we propose an approach for classification of medical images based on their modalities. We apply 2D adaptive noise-removal, median filter and 2D order-statistic filtering to remove noises and enhance the content of medical images. We propose to extract the features based on texture, appearance and signal. The extracted features are concatenated to each other. Different combinations of the extracted features are obtained for classification. KNN and SVM classifiers are then applied to classify medical imaging modalities. The stages of the proposed approach are shown in Fig. 1.

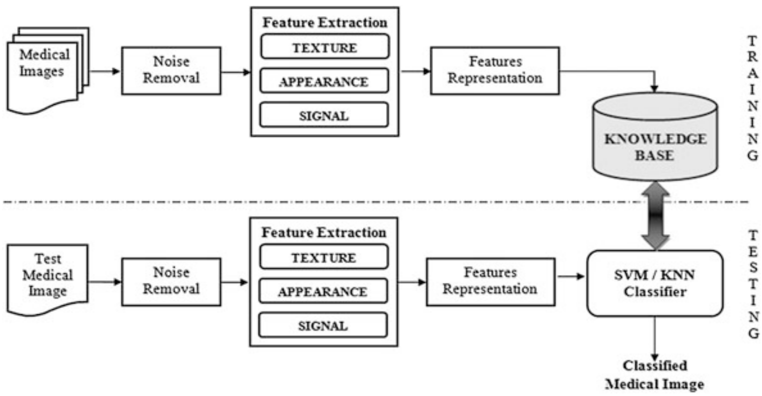


Fig. 1 Block-diagram of the proposed classification of medical imaging modalities approach

2.1 Feature Extraction

Texture, appearance, and signal features are extracted from the preprocessed medical images. In the following, discuss the details of feature extraction proposed in the paper.

2.1.1 Features Based on Texture

Different medical images are can be characterized by texture features of images. In this order, it is proposed to extract Local Binary Pattern (LBP) [9], Log-gabor [10], Discrete Cosine Transform (DCT) [11], Gray Level Difference Method (GLDM) [12], Gray-Level Run-Length Method (GLRM) [13], and Gray Level Co-occurrence Matrix (GLCM) as texture features. LBP is used to describe the gray-scale local texture of images. Log-gabor is a texture feature localizing the frequency information of the images. DCT is used to represent the transformation of the images in frequency domain. Various local statistical properties of medical images are computed by applying GLDM. To compute the number of gray level runs of various length of the medical images, GLRM is applied. GLCM is used to obtain the information about the position of pixels having similar gray level values in the images. The extracted features are concatenated to each other. Subsequently, the medical images are represented by the sets of texture feature vectors, each set represent the texture features of one image.

2.1.2 Features Based on Appearance

The appearances of medical imaging modalities are different due to the types of the devices capturing the medical images. Hence, appearance features can be considered as another feature that can represent different medical imaging modalities. The intensity values of medical images are considered as an appearance feature to represent the appearance of images. However, the sizes of the images are large. Instead of storing such huge intensity values, and in order to represent the appearance features of different medical imaging modalities in a uniform format, we apply Principal Component Analysis (PCA) on the medical images. After applying PCA, medical images are transformed into another space with dimensionality reduction. Hence, medical images are represented by such transformed appearance features. The obtained representation vectors are considered as appearance feature vectors for the images [14].

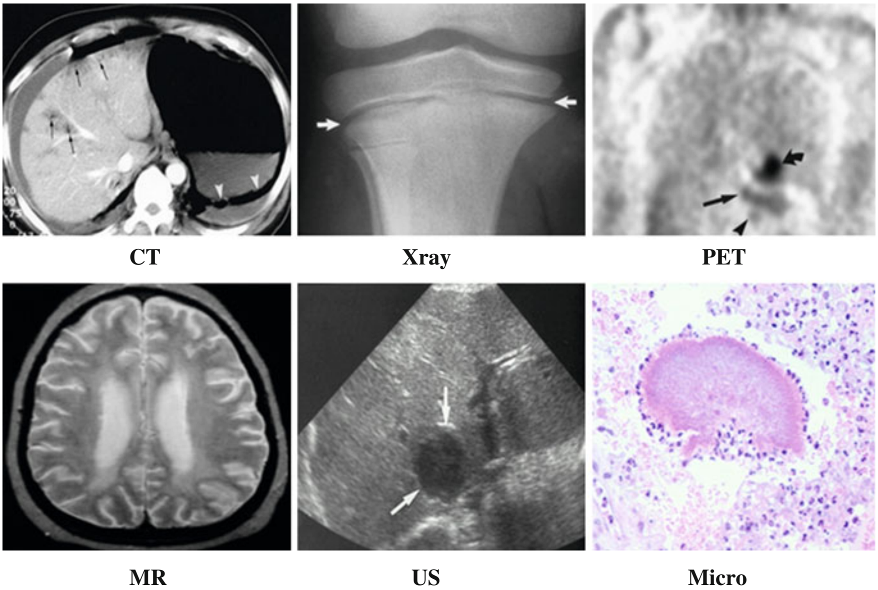


Fig. 2 Example samples of different medical imaging modalities of ImageCLEF2010

2.1.3 Features Based on Signal

Signal feature refers to the transformation of the input image into a set of representative signal features. We propose to extract Auto Regressive coefficient (AR), Mean Absolute Value (MAV), and Root Mean Square (RMS) as signal features [15]. AR coefficient represents the signal features based on frequency domain. MAV and RMS are extracted as another signal features that represent the signal features based on time domain. So, the extracted signal features are concatenated to each other for representing the signal features of an image.

2.2 Classification

Medical imaging modalities are labeled based on the domain knowledge of classes defined by the users. The represented features are fed to the K Nearest Neighbor (KNN) and Support Vector Machine (SVM) classifiers in order to classify medical imaging modalities.

Table 1 Classification accuracies using KNN classifier with different sets of experiments

	KNN classifier											
	Experiment sets											
	40:60				50:50				60:40			
	k = 1	k = 3	k = 5	k = 7	k = 1	k = 3	k = 5	k = 7	k = 1	k = 3	k = 5	k = 7
Texture	85.9	85.5	85.0	84.5	87.3	85.5	85.2	84.6	87.1	86.9	86.4	84.8
Appearance	88.2	88.1	87.4	87.1	88.9	88.8	88.7	87.1	87.1	88.9	88.5	88.4
Signal	78.6	78.4	77.7	77.6	79.5	79.4	78.7	78.5	89.5	80.4	79.5	78.8
Texture–appearance	87.5	86.4	85.7	84.1	87.1	86.9	86.5	86.1	80.7	87.3	87.0	86.6
Texture–signal	87.3	85.9	85.8	85.5	87.7	86.6	85.6	85.7	88.9	87.7	85.9	86.7
Appearance–Signal	81.7	81.5	81.1	80.6	82.4	82.2	82.2	81.7	83.7	83.4	82.2	82.0
Texture–appearance–signal	86.6	86.2	85.9	85.8	88.9	88.6	88.2	86.2	87.3	87.1	86.3	86.1

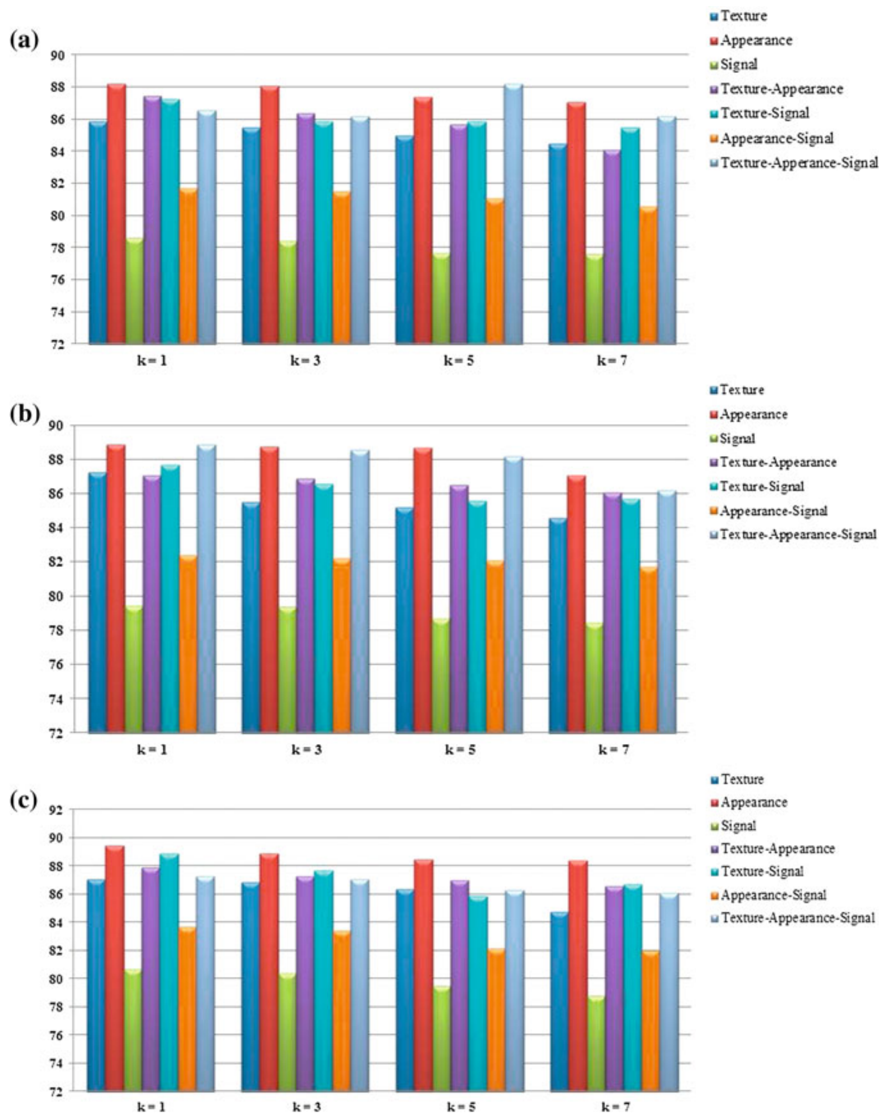


Fig. 3 Classification accuracy for classification of medical imaging modalities using KNN classifier under different sets of experiments (**a**) experiment set 1 (40:60), (**b**) experiment set 2 (50:50), (**c**) experiment set 3 (60:40)

3 Experimentation

In this section, we present the details of the benchmark database used to evaluate the performance of the proposed classification model along with the experimental results.

Table 2 Classification accuracies using SVM classifier with different sets of experiments

	SVM classifier		
	Experiment sets		
	40:60	50:50	60:40
Texture	93.15	93.82	94.18
Appearance	77.35	77.47	77.85
Signal	76.40	77.42	77.56
Texture–appearance	94.15	95.38	95.39
Texture–signal	93.55	94.74	95.05
Appearance–signal	82.45	83.14	83.27
Texture–appearance–signal	94.79	95.33	95.39

3.1 Image Dataset

IMageCLEF2010 is used to evaluate the performance of the proposed classification model. IMageCLEF2010 contains different medical imaging modalities. In this experimentation, we conducted experiments on classification of medical imaging modalities namely XR, MR, CT, PET, Ultrasound and Micro. Figure 2 shows examples of different medical imaging modalities samples of IMageCLEF2010 used in this work. The system has been randomly trained by selecting the 40, 50 and 60 % of the image samples of each class of the dataset. The selected random images have been considered as train samples and the remaining images as test samples. Hence, the system is evaluated in 3 different sets of experiment, i.e., (40:60), (50:50), and (60:40).

3.2 Experimental Results

To evaluate the performance of the proposed approach, different combination of the extracted features are studied and evaluated. In the first set of combinations each extracted feature (texture, appearance, signal) are considered individually. In the second set of combinations, any two of the extracted features are considered. Combinations of all the extracted features are considered as the third set of combinations.

KNN classifier has been used under varying k, for k = 1, 3, 5, and 7. Further, the SVM classifier is exploited by selecting the appropriate parameters.

The system has been trained with different sets of experiments, i.e., experiment set 1 (40:60), experiment set 2 (50:50), and experiment set 3 (60:40). The classification accuracy is calculated using Eq. 1.

$$\text{Accuracy Rate} = \frac{\text{Number of Correctly Classified Testing Samples}}{\text{Total Number of Testing Samples}} \times 100 \quad (1)$$

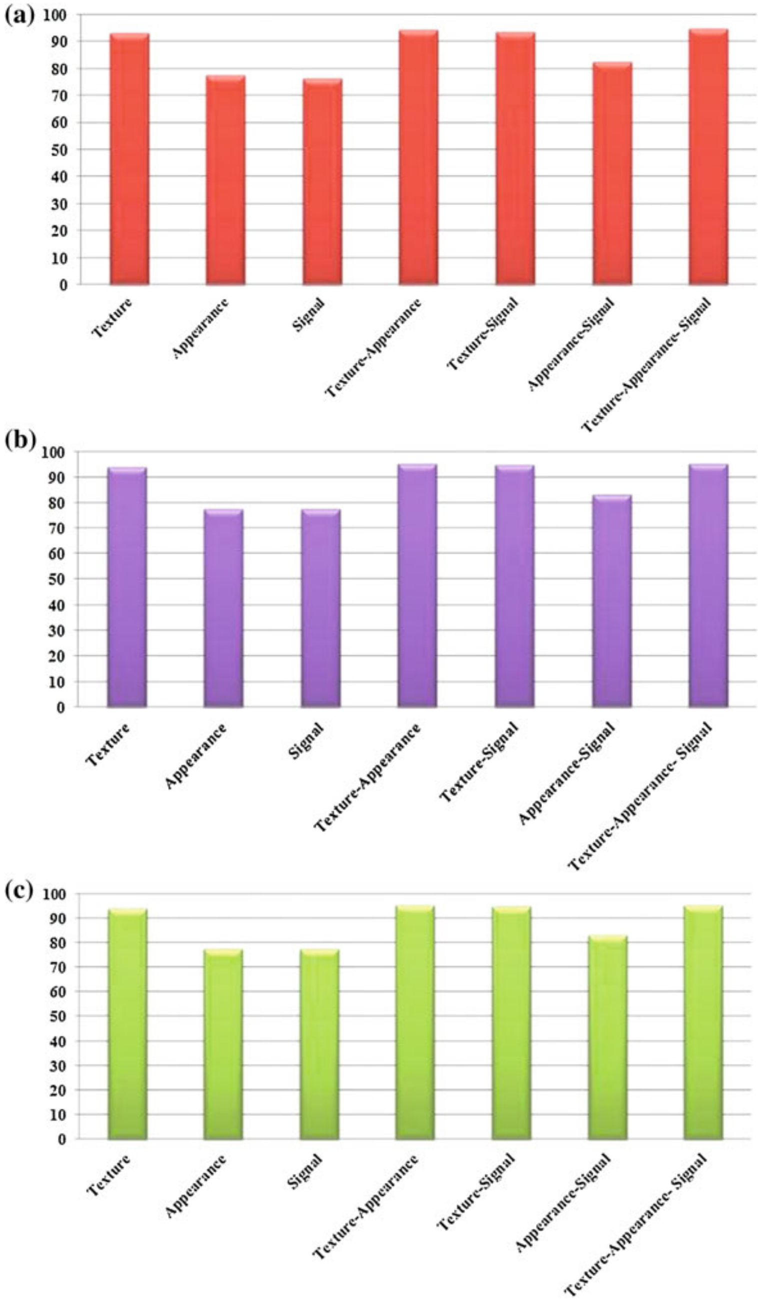


Fig. 4 Classification accuracy for classification of medical imaging modalities using SVM classifier under different sets of experiments (**a**) experiment set 1 (40:60), (**b**) experiment set 2 (50:50), (**c**) experiment set 3 (60:40)

Table 3 Qualitative comparison with the other existing methods for classification of medical imaging modalities

Methods	Feature extraction	Classifier	Dataset	Classification accuracy (%)
Pauly et al. [8]	Histograms Log-gabor Phase congruency	SVM/ Random Forest	IMage CLEF2010	93.53
Han and Chen [6]	Histogram descriptor Color intensity SIFT histogram Vocabulary words of image captions	SVM	IMage CLEF2010	93.89
Proposed approach	Texture Appearance Signal	KNN/SVM	IMage CLEF2010	95.35

The results of the classification accuracies using KNN classifier under different combination of features and different sets of experiments are tabulated in Table 1 and shown in Fig. 3. Table 2 presents the classification accuracies using SVM classifier under different combination of features and different sets of experiments. The results are as well shown in Fig. 4.

From Tables 1 and 2, it can be observed the highest classification accuracy is obtained using SVM classifier. The highest classification accuracy achieved 95.39 %.

We have also carried out a qualitative comparative analysis of our proposed approach with similar existing approaches in literature. The comparative analysis of the proposed approach with the recent works is tabulated in Table 3.

4 Conclusion

In this paper, 2D adaptive noise removal, median and 2D order-statistic filtering are applied to remove the noises and enhance the content of medical images. Texture, appearance, and signal features are extracted. Different combinations of the extracted features are considered for classification. KNN and SVM classifiers are applied to classify medical imaging modalities. The experimental results reveal the efficiency of our proposed approach.

Acknowledgments The authors would like to thank TM Lehmann, Department of Medical Informatics, RWTH Aachen, Germany, for making the database available for the experiments.

References

1. Lehman TM, Guld MO, Thies C, Fischer B, Keysers M, Kohnen D, Schubert H, Wein BB (2003) Content-based image retrieval in medical applications for picture archiving and communication systems. In: Proceeding of SPIE conference on medical imaging, vol 5033, pp 440–451
2. Mahmudur R Md, Bhattacharya P, Desai BC (2009) A unified image retrieval framework on local visual and semantic concept-based feature spaces. J Vis Commun Image Represent 20(7):450–462
3. Florea FI, Rogozan A, Bensrhair A, Dacher J-N, Darmoni S (2005) Modality categorization by textual annotations interpretation in medical imaging. Med Inf Bio-inf, pp 1270–1274
4. Malik A, Zremic T (2005) Classification of medical images using energy information obtained from wavelet transform for medical image retrieval. In: Proceedings of 7th international workshop on enterprise networking and computing in healthcare industry, pp 124–129
5. Florea F, Barbu E, Rogozan A, Bensrhair A, Buzuloiu V (2006) Medical image categorization using a texture based symbolic description. In: Proceeding of international conference on image processing, pp 1489–1492
6. Han X, Chen Y (2010) ImageCLEF2010 modality classification in medical image retrieval: multiple feature fusion with normalized kernel function. In: Image CLEF2010 workshop
7. Kalpathy CJ, Hersha W (2007) Automatic image modality based classification and annotation to improve medical image retrieval, MEDINFO, pp 1334–1338
8. Pauly O, Mateus D, Navab N (2010) ImageCLEF2010 working notes on the modality classification subtask, cross language image retrieval workshop
9. 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(7):971–987
10. Fischer S, Sroubek F, Perrinet L, Redondo R, Cristo′ Bal G (2007) Self-invertible 2D log-gabor wavelets. Int J Comput Vision 75(2):231–246
11. Rajakumar K, Muttan S (2010) Medical image retrieval using modified DCT. In: Procedia Computer Science, ICEBT2010, vol 2, pp 298–302
12. Weszka JS, Dyer CR, Rosenfeld A (1976) A comparative study of texture measures for terrain classification. IEEE Trans Syst, Man, Cybern, SMC(6):269–285
13. Galloway MM (1975) Texture analysis using gray level run lengths. Computer Graph Image Process 4:172–179
14. Imran M, Rao A, Kumar GH (2010) Multibiometric systems: a comparative study of multi-algorithmic and multimodal approaches. Procedia Computer Science, ICEBT2010, vol 2, pp 207–212
15. Phinyomark A, Limsakul C, Phukpattaranont P (2009) A novel feature extraction for robust EMG pattern recognition. J Comput 1(1):71–80

Experimental Verification of Squirrel Cage Induction Motor Using Current Signature and Virtual Instrumentation Topology

K. Vinoth Kumar, S. Suresh Kumar and S. Daison Stallon

Abstract Three phase squirrel cage induction motors are workhorses of industry and are the widely used industrial drives. With the passage of time the machine may develop faults and may hamper the production line that may lead to production and financial losses. A proper planning of maintenance schedule and condition monitoring is essential to reduce such financial loss and shut down time. A condition monitoring system, which can predict and identify the prefault condition, is the need of the age to prevent such unwanted breakdown time. The MCSA (Motor Current Signature Analysis) technique is found one of the most frequently used technique to identify the prefault condition. This paper focuses on experimental results to prove that MCSA Technique can identify the good and cracked rotor bar in three phase squirrel cage induction motors under no-load and different load conditions and also simulated in Virtual Instrumentation. The diagnostics strategy is presented in this paper and variables that influence the diagnosis are discussed.

Keywords Induction motor · Virtual instrumentation · Broken bars

K. Vinoth Kumar (✉) · S. Daison Stallon
Department of Electrical and Electronics Engineering,
School of Electrical Sciences, Karunya University,
Coimbatore 641114, Tamilnadu, India
e-mail: kvinoth_kumar84@yahoo.in

S. Suresh Kumar
Department of Electronics and Communication Engineering,
Dr. N.G.P.Institute of Technology,
Coimbatore 641048, Tamilnadu, India

Mohan S. and S. S. Kumar (eds.), *Proceedings of the Fourth International Conference on Signal and Image Processing 2012 (ICSIP 2012)*, Lecture Notes in Electrical Engineering 222,
DOI: 10.1007/978-81-322-1000-9_45,

1 Introduction

Squirrel Cage Induction motor have been so far considered as most suitable industrial drive. Machine is rugged, safe and simple in operation and maintenance. These machines are designed for a specific lifetime, but it is not a proven fault development machine. Due to aging different faults may develop and may affect the production loss leading to financial loss [1]. Maintenance schedule keeps the machine in working condition. Many Electrical components, which are susceptible to failures, are not under the preview of routine maintenance schedule. Such unwanted machine shut down cost both time and money that could be avoided if an early warning system is available against impending failures.

Fault diagnosis of electrical machines can lead to greater plant availability, extended plant life, higher quality products, and smoother plant operations. Numerous fault detection methods have been proposed to identify the faults [2]. One of the most frequently used fault detection method is Motor Current Signature Analysis (MCSA). MCSA is a noninvasive, On-Line monitoring Technique for the diagnosis of faults in induction motors such as broken rotor bars, abnormal levels of air gap eccentricity, shorted turns in low voltage stator windings, and certain mechanical problems [3, 4]. This paper establish through focuses on experimental results that motor current signature analysis (MCSA) can diagnose broken rotor bar fault in three phase squirrel cage induction motors. This technique implemented on three phase, 3 hp, 415 V, 50 Hz, Squirrel cage induction motor with different loading conditions with single and double crack in the rotor bar. The Results obtained simply qualifies the theory and establish the importance of MCSA in diagnose the rotor bar faults.

2 Induction Motor Fault Detection Using MCSA

Motor Current Signature Analysis technique [5] applied to 3 hp, three phase 415 V, 50 Hz, squirrel cage induction motor for various load conditions. A schematic circuit diagram and a theoretical current spectrum for an idealized machine broken rotor bar are shown in Figs. 1 and 2 respectively.

Usually a decibel (dB) versus frequency spectrum is used in order to give a wide dynamic range and to detect the unique current signature patterns that are characteristic of different faults. To detect the broken rotor bar, motor current signal spectrum is observed in terms of Amplitude in decibel (dB) versus frequency spectrum in Hz and by measuring the amplitude difference between line frequency amplitude and the amplitude of the first pole pass sideband below line frequency. Then with reference to IEEE survey [6–8] is shown in Table 1, the condition of the rotor is identified. Both machine condition as well as lubricant condition too [2]. In such cases, lubricating oil may be used as a diagnostic medium that carries the wear particles generated from the surfaces. But doing analysis program of wear debris, the

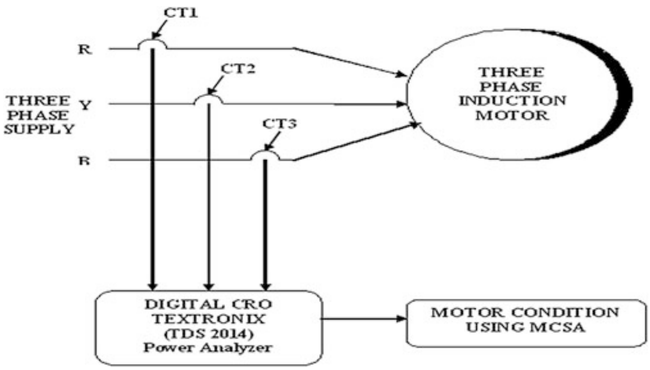


Fig 1. MCSA arrangement in induction motor

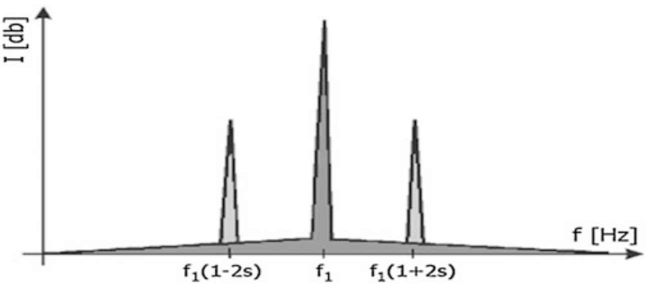


Fig 2. The idealized current spectrum

Table 1 Identification of rotor condition using MCSA

Amplitude difference dB	Rotor condition (With at least 75 % of rated load)
>60	Excellent
54–60	Good
48–54	Moderate
42–48	Bar crack may be developing or high resistance joints
36–42	Two bars may be cracked or high resistance joints likely
30–36	Multiple cracked or open bars or end ring probable
<30	Multiple broken bars and/or end rings very likely

condition of the machine parts can be judged properly. The condition-monitoring program based on oil analysis has been updated over the past few decades. The most important factor is aging of machinery in machine diagnostics as it involves machine depreciation. The rates of change of machine components always depend upon the measurable parameters, where maintenance of lubricant plays a vital role, which has been shown in Fig. 1. This kind of system can overcome all possible hazards at its optimum level so that a wider range of machine condition monitoring can be followed up by utilizing wear particle analysis [1].

In order to identify the rotor condition of three-phase squirrel cage induction motor using MCSA a modern laboratory test bench was set up at our VIT Laboratory. The rated data of the tested three-phase squirrel cage induction machine were: 2.2 KW Three phase squirrel cage induction motor 400 V, 50 Hz, 4 Pole, and Nr = 1475 rpm, $I_{nl} = 2.5$ A, $I_{fl} = 4.5$ A delta connection. It consists of two coupled electrical machines: the machine to be tested and a DC machine connected with loading rheostat. Fluke current transducers are used to measure the currents. The measured current signals were processed using the Fast Fourier Transformation (FFT) with Tektronics TDS 2014 Power Analyzer. Tests were carried out for seven different loads with the healthy rotor, and with similar motors having single cracked rotor bar and double cracked rotor bar. The rotor faults were provoked interrupting the rotor bars by one bar cracked and two bars cracked into the rotor see in Figs. 3 and 4. The results obtained for the healthy motor and those having rotor faults were compared.

Fig. 3 One bar cracked rotor

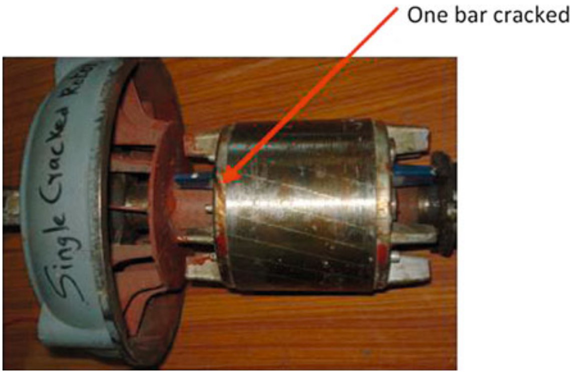


Fig. 4 Two bar cracked rotor

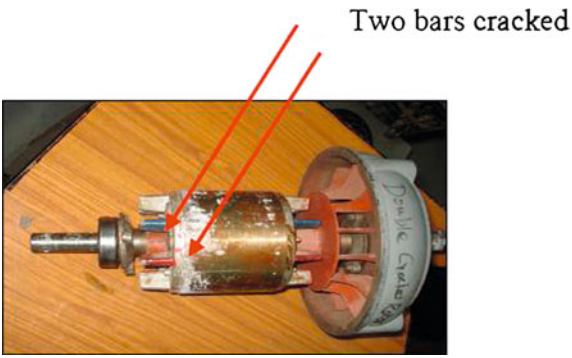


Fig. 5 Stator current spectrum for two bar cracked rotor bar

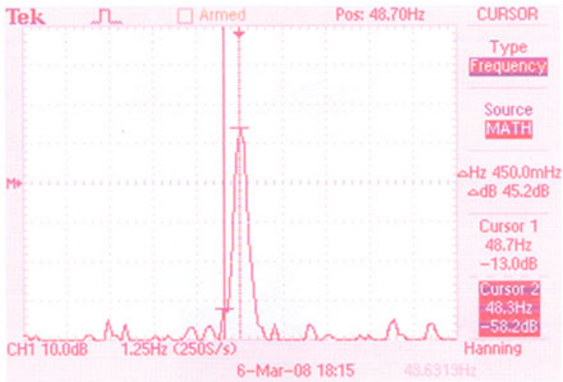


Fig. 6 Stator current spectrum for one bars cracked rotor bar

3 Results and Discussion

3.1 With No Load Condition

Here the stator current FFT spectrum for induction motor running under no-load condition with good rotor, one bar cracked rotor and two bar cracked rotor are shown below in Figs. 5 and 6 respectively. In the Fig. 5, FFT spectrum there is no sidebands are presented across the supply frequency and in Fig. 6 slit sidebands are presented across the supply frequency. This result satisfies the IEEE survey table for rotor condition Identification shown in Table 2.

3.2 With Load Condition

Here the stator current FFT spectrum for induction motor running under load condition 2.8 A with good rotor, one bar Cracked rotor and two bar cracked rotor

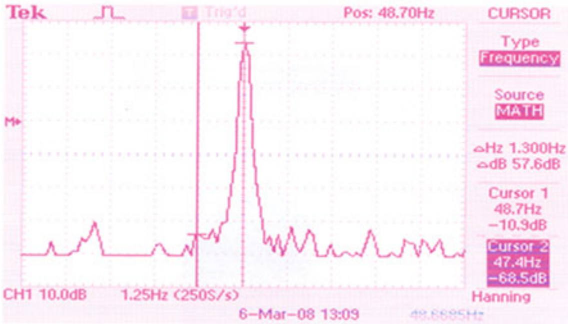


Fig. 7 Stator current spectrum for good rotor at 2.8 A

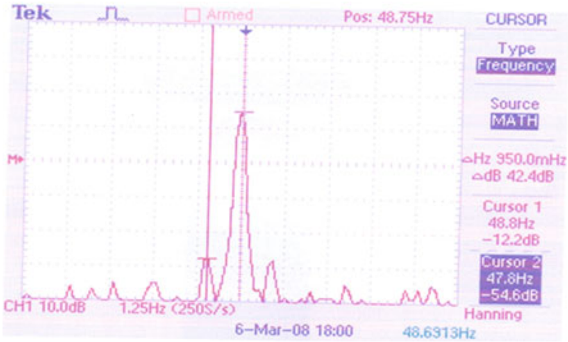


Fig. 8 Stator current spectrum for one cracked rotor at 2.8 A

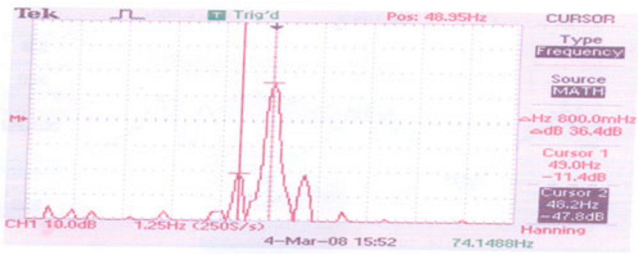


Fig. 9 stator current spectrum for two bar cracked rotor for at 2.95 A

are shown in Figs. 7, 8 and 9 respectively. In the Fig. 7, FFT spectrum there is no Sidebands are presented across the supply frequency and in Figs. 8 and 9 the sidebands are presented across the supply frequency. These results satisfies the IEEE survey table for Rotor condition Identification shown in Table 3.

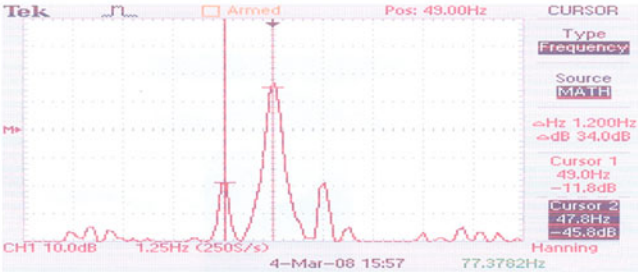


Fig. 10 Stator current spectrum for two bar cracked rotor—load case

Here the stator current FFT spectrum for induction motor running under load condition 2.8 A with good rotor, one bar cracked rotor and two bar cracked rotor are shown in Fig 10 respectively. In the Fig. 10, FFT spectrum there is no sidebands are presented across the supply frequency (Tables 2 and 3).

4 Simulation of Induction Motor Using Virtual Instrumentation

The three-phase induction motors also operate on the principle of a rotating magnetic field. As already stated, the rotating magnetic field has a constant magnitude rotating at synchronous speed defined by the supply frequency.

Table 2 No load case for induction motor

Parameters No load	Good condition rotor	One bar cracked rotor	Two bar cracked rotor
f_s	48.60 Hz	48.70 Hz	49.2 Hz
N_s	1458 rpm	1461 rpm	1476 rpm
N_r	1449 rpm	1451 rpm	1467 rpm
Slip s	0.006172	0.006844	0.00609
Sideband±2sfs	±0.6 Hz	±0.66 Hz	±0.599 Hz
f_s-2sf_s in Hz	48 Hz	48.3 Hz	48.50 Hz
f_s+2sf_s in Hz	49.2 Hz	49.366 Hz	49.799 Hz
f_s-2sf_s in dB	−66.9 dB	−58.2 dB	−64.2 dB
f_s in dB	−13.7 dB	−13.0 dB	−27.4 dB
Amplitude difference $f_s-(f_s-2sf_s)$ in dB	53.2 dB	45.2 dB	36.8 dB

Table 3 Load case for induction motor

Parameters	Good condition rotor 2.8 A	One cracked rotor bar 2.5 A	Two cracked rotor two cracked rotor bar 2.95 A bar 2.95
f_s	48.70 Hz	48.75 Hz	48.95 Hz
N_s	1461 rpm	1462.5 rpm	1468.5 rpm
N_r	1439 rpm	1440 rpm	1455 rpm
Slip s	0.015058	0.01538	0.009193
Sideband±2sfs	1.466	1.499	0.899
f_s-2sf_s in Hz	47.4 Hz	47.8 Hz	48.05 Hz
f_s+2sf_s in Hz	50 166 Hz	50.2495 Hz	49.849 Hz
f_s-2sf_s in dB	−68.5 dB	−54.6 dB	−47.8 dB
f_s in dB	−10.9 dB	−12.2 dB	−11.4 dB
Amplitude difference $f_s-(f_s-2sf_s)$ in dB	57.6 dB	42.4 dB	36.4 dB

where the coefficients are

$$
\begin{aligned}
k_1 &= \Delta t f'(x_{n'} y_n) \\
k_2 &= \Delta t f'(x_n + \Delta t/2, y_n + k_1/2) \\
k_3 &= \Delta t f'(x_n + \Delta t/2, y_n + k_2/2) \\
k_4 &= \Delta t f'(x_n + \Delta t, y_n + k_3)
\end{aligned}
$$

Here n is the time step, $\Delta t = t_{n+1} - t_{n'} f'(x_{n'} y_n) \;=\; dy_n/dx_{n'}$

The simulation also provides an inverse transformation to determine the abc reference frame that corresponds to the real parameters of the motor for easy comparison. Conversion to the abc reference frame is achieved by using the following transformations. The front panel of the VI is shown in Fig. 11. Before starting the simulation, the user should enter the motor and the load parameters, and set the time step. Alter the simulation is instigated, the graphs display the estimated values of the electromagnetic torque and rotor speed. However, make sure that the motor parameters you entered are practical. As a guide, the sample settings in Table 4 which were obtained using a real machine in our laboratory,

Table 4 Simulation parameters of induction motor

Number of Poles: 4	$X_m = 43\ \Omega$
$R_s = 1.05\ \Omega$	Speed, $\omega_m = 157$ rad/s
$R_r = 0.16\ \Omega$	$V_{ds} = 0$ V
$X_{ls} = 2.4\ \Omega$	$V_{qs} = 225.2$ V
$X_{lr} = 0.37\ \Omega$	J = 0.39 kg m2
F = 50 Hz	$\Delta t = 0.01$ s
$T_L = 30$ Nm	

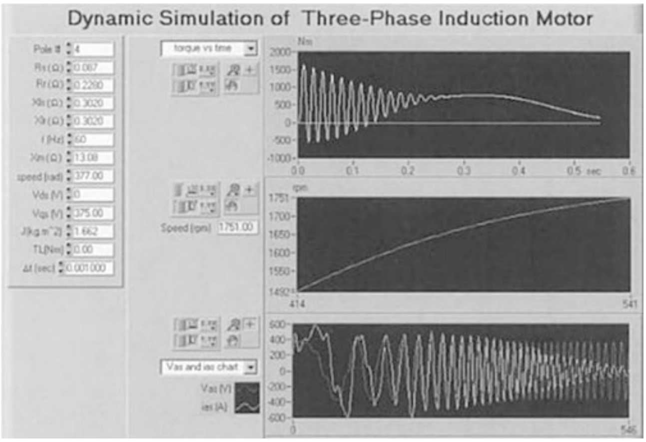


Fig. 11 Simulation of induction motor using LabVIEW

can be used. The tests performed in Fig. 11 can provide all the equivalent circuit parameters and the total moment of inertia of the rotating system.

5 Conclusion

This paper presented the experimental analysis on identification of three-phase squirrel cage induction motor rotor condition by using MCSA method in Good and Cracked rotor/broken rotor condition and also in Virtual Instrumentation. Also it stated that, as the load is increased, the magnitude of the sideband frequency components is also decreased. These results have clearly demonstrated that MCSA is a powerful technique for monitoring the health of three-phase induction motor rotor.

References

1. Thomson WT (1999) A review of on-line condition monitoring techniques for three-phase squirrel cage induction motors. The Robert Gorden University, Scotland, IEEE
2. Zhongming YE, Bin WU (2000) A review on induction motor online fault diagnosis. Ryerson Polytechnic University, Canada, IEEE
3. Saravana Kumar R, Ray KK, Swaminathan G (2004) Fuzzy logic based fault identification system for a three phase induction motor. In: Proceedings of the 28th NSC-2004 on 16–18
4. Thomson WT (2001) On-line MCSA to diagnose shorted turns in low voltage stator windings of 3-phase induction motors prior to failure, IEEE, PES&IAS IEMDC. MIT, Boston
5. Thomson WT, Fenger M (2001) Current signature analysis to detect induction motor faults. IEEE Trans IAS Mag 7(4):26–34

6. Bishop T (2003) Squirrel cage rotor testing. EASA Convention 2003, Moscone convention Centre, San Francisco, CA, June 30
7. Tavner P, Penman J (2011) Condition monitoring of electrical machines, Research Studies Ltd. Wiley, London
8. Vas P (1995) Parameter estimation, condition monitoring & diagnosis of electrical machines. Peter vas, Oxford University Press

An Evaluation of Edge Detection Algorithms for Mammographic Calcifications

Vikrant Bhateja, Swapna Devi and Shabana Urooj

Abstract Edge detection is an important module in medical imaging for diagnostic detection and extraction of features. The main limitation of the existing evaluation measures for edge detection algorithms is the requirement of a reference image for comparison. Thus, it becomes difficult to assess the performance of edge detection algorithms in case of mammographic features. This paper presents a new version of reconstruction estimation function for objective evaluation of edge enhanced mammograms containing microcalcifications. It is a non-reference approach helpful in selection of most appropriate algorithm for edge enhancement of microcalcifications and also plays a key role in selecting parameters for performance optimization of these algorithms. Simulations are performed on mammograms from MIAS database with different category of background tissues; the obtained results validate the efficiency of the proposed measure in precise assessment of mammograms (edge-maps) in accordance with the subjectivity of human evaluation.

Keywords Edge detection · MIAS · Microcalcifications · Non-reference · Reconstruction estimation function

V. Bhateja (✉)
Department of Electronics and Communication Engineering, SRMGPC,
Lucknow, Uttar Pradesh, India
e-mail: bhateja.vikrant@gmail.com

S. Devi
Department of Electronics and Communication Engineering,
NITTTR, Chandigarh, India
e-mail: swapna_devi_p@yahoo.co.in

S. Urooj
Electrical Engineering Department, School of Engineering, GBU, Gr. Noida,
Uttar Pradesh, India
e-mail: shabanabilal@gmail.com

Mohan S. and S. S. Kumar (eds.), *Proceedings of the Fourth International Conference on Signal and Image Processing 2012 (ICSIP 2012)*, Lecture Notes in Electrical Engineering 222,
DOI: 10.1007/978-81-322-1000-9_46,

1 Introduction

Microcalcifications are tiny specks of calcium deposits that can be scattered throughout the mammary gland, or occur in clusters. They vary in size, shape, signal intensity and contrast, and hence their classification as benign or malignant requires accurate preservation of their edges and other morphological details. Detection of microcalcifications will therefore require an explicit representation of edges by effective coding in the high frequency region of the mammogram. Hence, accurate detection of microcalcifications becomes a priority in the diagnosis of mammographic images. Selection of a sensitive edge detector in medical imaging helps not only in extraction of features for diagnostic detection but also aids in deciding the future therapies and treatment patterns [1]. Conventional gradient based edge detectors (Sobel, Prewitt etc.) are known for their simplicity of operation but yield inaccurate results and lack sensitivity to noise. Zero crossing edge detectors (Laplacian) produces detection of only those edges possessing fixed characteristics in all the directions but fail to perform in the presence of noise. LoG edge detectors fail to perform at locations where there is a variation in gray levels, (like corners and curves), therefore are not frequently used in machine vision. Gaussian edge detectors (Canny) tend to improve the signal-to-noise ratio by smoothening the image, yielding better detection of edges in presence of noise. However, these operators fail to yield optimal results using a fixed operator [2–4]. A common approach for the calcification detection task performs localization of high spatial frequencies in the digital mammograms using wavelet transform [5]. Other non-wavelet based methods try to make maximum use of the fact that calcifications have much higher intensity values than the surrounding breast tissues [6]. CLAHE [7] provides significantly sharp and defined edges but, these are coupled with the enhancement of other unwanted information. Variants of Unsharp Masking (UM) [8–10] did not produced satisfactory enhancement of edges in mammograms on account of the presence of some over shoots in the region of interest (ROI). Stojic et al. [11, 12]proposed an algorithm for local contrast and edge enhancement of calcifications using morphological top-hat and bottom-hat operators. However, generation of a difference image by this algorithm is an irreversible operation leading to an output image with negative pixel values. Domínguez et al. [13] proposed a method for mammographic image feature extraction, which generated the edge-map (of calcifications) using Co-ordinate Logic Filters (CLF) with a rhombic structuring element of order 3×3. Another related work by Santhaiah et al. [14] used morphological gradient based operator for detection of edges in medical images. Bhateja and Devi [15] proposed an edge detection algorithm for microcalcifications using a two-stage morphological filter. They used variable sized structuring elements at two stages to trap the microcalcification clusters as well as tiny scattered microcalcifications. From the above discussion, it can be ascertained that edge-maps (containing the segmented calcifications) are binary images obtained after thresholding the mammographic images obtained from edge enhancement algorithms. Many techniques have been proposed in literature for performance evaluation of these edge detection algorithms but the majority is based on particular criteria and constraints that limit their usage in a generalized mode. Hence, there is a requirement to devise an automated as well as precise metric for quantitative evaluation of mammographic edge maps and also be used as a tool for predicting and benchmarking the performance of various edge detection algorithms for mammograms. It is also necessary that this mathematical assessment must be consistent to that of subjective assessment by human observation, as these images are ultimately analyzed by medical experts for diagnosis. The remainder of this paper is structured as follows: Sect. 2 details the classification of measures for quality evaluation of edge-maps. The proposed version of edge-map evaluation procedure is described in Sect. 3. The results and discussions are given under Sects. 4 and 5 draws the conclusion.

2 Quality Evaluation Measures for Edge Maps

Edge evaluation methods can be broadly modeled as *subjective* and *objective* in nature. Subjective evaluation measures are generally based on visual comparison of the obtained edge maps to that of original inputs by human observers. An edge enhancement system analyzed under 'N' different viewing conditions evaluates to a set of N discrete estimation results. These could be appropriate methods for estimating the quality specifically in medical imaging which require the end results to be viewed by expert radiologists. However, these measures are very much dependent on observers' subjectivity and are generally inconvenient, time consuming and expensive [16, 17].

The versatility of objective evaluation methods for edge maps, especially for medical images is governed by the fact that it can dynamically monitor the quality independent of any viewing condition and would aid the radiologists in accurate diagnosis, thereby deciding future therapies and treatment patterns [18]. Objective evaluation measures can be further categorized as Reference-based and Non-reference evaluation methods.

2.1 Reference Based Evaluation Methods

These are signal fidelity measures which compares the transformed edge map to original edge map (consisting of true edges, also referred to as the ground truth image) by providing a quantitative score that gives an account of the degree of similarity between them. These measures can conveniently estimate the magnitude of error but do not depicts a visualization of edge reconstruction quality. Full-reference methods, like Pratt's Figure-of-Merit (PFOM) [19, 20], find application for synthetic images; where there is a complete prior knowledge of true edge locations (known ground truth image is available) but fail to assess for real (natural) images. Reduced-reference evaluation methods for edge maps tend to generate an estimated edge map [21] of the original image for comparison and calculation of degree of similarity with the output edge map. Edge Similarity Index (ESI) [22] is one example of this category of measures, where the statistical properties of the edge maps are used for estimating the quality of edge maps. These measures cease to give satisfactory results in noisy environments.

2.2 Non-Reference Evaluation Methods

Non-reference evaluation methods quantify the edge detection performance without any prior knowledge of the true edge pixels or their locations. Reconstruction estimation functions [23] are non-reference measures for evaluation of output edge maps for real images. These functions operate by deriving a reconstructed estimation of the original input image from the output edge map, based on the fact that edges contain the most important information of in any image. The reliability of reconstruction estimation function is questionable as during assessments, the image reconstruction is mainly performed considering primarily the edge pixels. Although, the non-edge pixels also carry a necessary component of information and hence their role in generation of reconstructed image cannot be ignored or given reduced significance. Thus, generation of interpolating values in case of non-edge pixels at times becomes crucial. This process becomes more significant in case of mammograms, as improper reconstruction even for non-edge pixels may result in loss of diagnostically useful information. In the Reconstruction estimation methods, the quality of the reconstructed image will be more pronounced if either there is an increase in the number of edge pixels or their density. However, this is not the case in mammograms, because the calcifications are either scattered throughout the region as tiny specks or occur in small clusters. Hence, direct application of Carlson's Reconstruction estimation functions [24] may not prove to be useful in assessment of mammographic edge maps. B. Govindarajan et al. [25] used the weighted average approach for interpolation of edge pixels during the reconstruction estimation. However, the approach was limited owing to its incapability to preserve sharpness and lacked satisfactory performance in noisy environments. As an improvement of averaging approach, the authors also proposed median based interpolation scheme [26] for reconstruction. The usage of median based interpolation blurred the finer details (microcalcifications) and was also limited in retaining sharpness upon reconstruction. Hence, this paper proposes a new version of reconstruction based estimation function for performance evaluation of edge detection algorithms for mammographic calcifications.

3 Proposed Evaluation Procedure for Edge-Maps (of Calcifications)

The quality evaluation procedure for mammogram edge-maps (containing calcifications) consists of three important modules, namely: Estimation of edge pixel information, Reconstruction Estimation Module followed by Error Assessment of the reconstructed image as shown in Fig. 1.

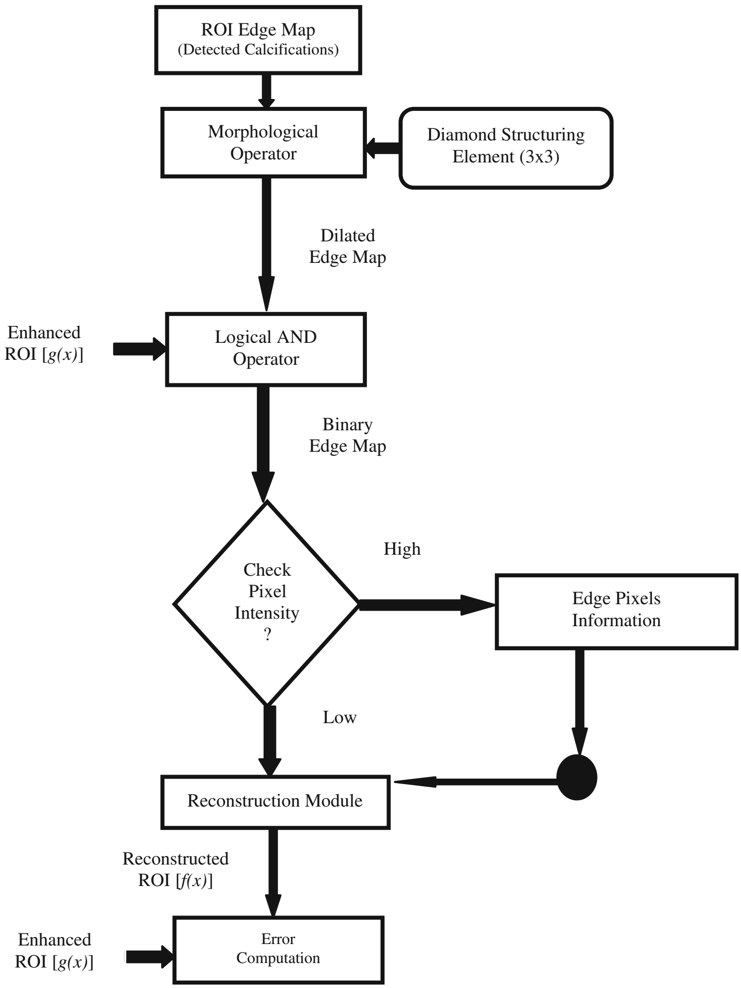


Fig. 1 Flow diagram of proposed evaluation method for edge-maps (calcifications)

3.1 Estimating Edge Pixel Information

The first step in the estimation of edge pixels can be initiated by morphological dilation of the output edge map using a flat (diamond shaped) structuring element of size 3 × 3. The dilated edge map is logically multiplied with the edge enhanced ROI containing calcifications (prior to thresholding). The binary ROI obtained as the product is checked for the edge pixel locations. A new ROI is reconstructed by locating the corresponding pixels in edge map with magnitude unity (defined as edge points of calcifications). This reconstructed ROI contains the estimated edge pixel information. In case, the binary ROI returns a non-edge pixel, then the interpolation pixel value in the reconstructed ROI is determined using the reconstruction module.

3.2 Reconstruction Module

The performance evaluation of the edge detection algorithms can be ascertained by filling in between contours of the obtained edge map, and reconstructing the original image. If the gray-level information in the binary edge-map, determines the edge pixels, then the remaining part of the ROI $f(x, y)$ (containing the non-edge pixels) is reconstructed by interpolation using the minimum variation criteria [27] given in (1).

$$\iint \left(\frac{\partial f}{\partial x}\right)^2 + \left(\frac{\partial f}{\partial y}\right)^2 dxdy \tag{1}$$

The above variation measure can be converted to the discrete version using Eq. (2).

Evaluating the above variation criteria as a constrained optimization problem; the process starts by computing the values of f minimizing above equation subject to the condition that f attains particular values along the edges.

$$Q = \sum_x \sum_y (f_{x,y} - f_{x,y-1})^2 + (f_{x,y} - f_{x-1,y})^2 \tag{2}$$

This can be further solved using the successive over-relaxation approach iteratively for finding the minima of the Eq. (2).

$$f^{(i+1)}(x,y) = f^{(i)}(x,y) + \frac{\phi}{4}\Delta f^{(i)}(x,y) \tag{3}$$

$$\begin{aligned} \text{where}: \Delta f^{(i)}(x,y) \\ &= f^{(i+1)}(x-1,y) + f^{(i+1)}(x,y-1) + f^{(i)}(x+1,y) + f^{(i)}(x,y+1) \\ &\quad - 4f^{(i)}(x,y) \end{aligned} \tag{4}$$

This approach is very much analogous to a shift variant low-pass filter applied iteratively to the entire ROI. Stable results can be obtained for relaxation parameter (ϕ) values lying between 0 and 2. Application of this approach tends to smoothen any of the unwanted gray-level variations or artifacts which might be introduced on account of incorrect pixels at non-edge points [27].

3.3 Error Computation

The error computation module, simply compares the reconstructed ROI [*f(x,y)*]with the original enhanced ROI [*g(x,y)*] to estimate the amount of error in the reconstructed ROI. For an ROI of size *M* x *N*, the Mean Squared Error (*MSE*) and Mean Absolute Error (*MAE*) is given as in (5) and (6) respectively. Lower the values of these error indices, implies that better is the quality reconstructed image. Hence, better is the performance of edge detection algorithm.

$$MSE = \frac{1}{M.N}\sum_{i=1}^{M}\sum_{j=1}^{N}[g(i,j) - f(i,j)]^{2} \tag{5}$$

$$MAE = \frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}|g(i,j) - f(i,j)| \tag{6}$$

4 Results and Discussion

4.1 Data Collection and Pre-processing

The data used in these experiments has been taken from the Mammographic Image Analysis Society (MIAS) database [28], which is publically available and one of the most easily accessed databases. The original input mammographic images read from the MIAS database are subjected to normalization. ROI (containing calcifications) of size 256 $\times$ 256 are extracted to eliminate the unwanted portion of the image that consists of most of the background area. Different edge detection algorithms are applied on 18 images, containing calcifications embedded in a background of fatty-tissues [S. No. (1)–(2)], fatty-glandular tissues [S. No. (3)–(6)], and dense-glandular tissues [S. No. (7)–(10)] to assess their versatility on various types of mammograms.

4.2 Simulation Results

For performance evaluation, the output edge maps obtained from the following edge detection algorithms are used: CLF [6, 13], is named as ED-I, Morphological gradient operator [14], named as ED-II and Morphological filtering using top-hat and bottom hat Transformations [12], named as ED-III, Two-stage iterative morphological filter [15], termed ED-IV.

These edge detection algorithms are applied to the enhanced ROI, which following by thresholding yields the output edge map, whose quality is evaluated using the proposed measure. It can be observed in the edge maps of Fig. 2d that some of the microcalcification clusters are missed owing to usage of small sized structuring element. In Fig. 2e the edge map tend to miss some of the tiny scattered calcifications, on the other hand Fig. 2f shows detection of very few calcifications. However, a reasonably good response can be seen in Fig. 2g which shows recovery of both the micro-calcification clusters as well as the scattered ones. It can be interpreted from Tables 1 and 2 that the values of error metrics *MSE* and *MAE* are least for ED-IV (as in Fig. 2g), validating its superior performance in comparison to other edge detectors. As visible in Fig. 2f, most of the calcifications are missed by the edge detector; hence ED-III yields the highest error values among all four edge maps. Nearly comparable error values are obtained for ED-I and ED-II respectively. Similarly, edge maps are obtained using the edge detectors ED-I to IV for another two test images mdb213 and mdb223 as shown in Figs. 3 and 4 respectively.

It can be again visualized that aligned as well as clustered calcifications are generally missed in the edge maps of Fig. 3d–f. As the approach in ED-IV uses a two stage edge detector; the first pass of ED-IV uses a structuring element (of size 5×5),

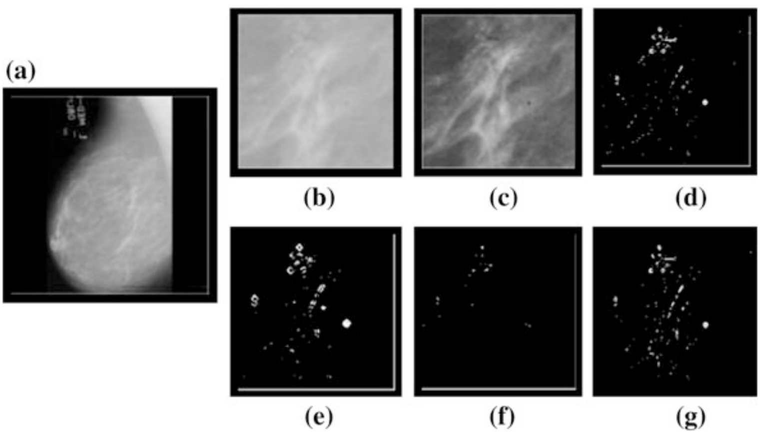


Fig. 2 **a** Original mammogram (mdb231) containing calcification embedded in a background of fatty-breast tissues. **b** Extracted ROI. **c** Enhanced ROI. Edge maps obtained using: (**d**) ED-I, (**e**) ED-II, (**f**) ED-III, (**g**) ED-IV

Table 1 Calculation of *MSE* ($\times 10^6$) for r different edge detection algorithms

S. No.	Ref. No.	ED-I	ED-II	ED-III	ED-IV
(1)	mdb231 (M)	9.8975	9.9975	10.197	9.4507
(2)	mdb238 (M)	9.7244	9.9234	10.045	9.0045
(3)	mdb211 (M)	10.903	8.4645	8.4146	6.5155
(4)	mdb213 (M)	9.2601	6.7692	6.8989	6.3067
(5)	mdb219 (B)	6.7966	6.4810	7.1550	5.3016
(6)	mdb233 (M)	10.757	8.8075	17.867	8.2424
(7)	mdb223 (B)	9.9993	9.8396	10.603	9.1446
(8)	mdb239 (M)	61.550	46.487	65.560	40.943
(9)	mdb240 (B)	65.121	66.030	96.521	56.982
(10)	mdb241 (M)	8.0183	7.8074	9.1274	7.4785

Table 2 Calculation of *MAE* for different edge detection algorithms

S. No.	Ref. No.	ED-I	ED-II	ED-III	ED-IV
(1)	mdb231 (M)	0.0056	0.0040	0.0082	0.0020
(2)	mdb238 (M)	0.0046	0.0042	0.0073	0.0019
(3)	mdb211 (M)	0.0058	0.0026	0.0070	0.0011
(4)	mdb213 (M)	0.0067	0.0039	0.0078	0.0017
(5)	mdb219 (B)	0.0050	0.0038	0.0068	0.0014
(6)	mdb233 (M)	0.0060	0.0030	0.0071	0.0014
(7)	mdb223 (B)	0.0070	0.0057	0.0107	0.0015
(8)	mdb239 (M)	0.0106	0.0070	0.0140	0.0045
(9)	mdb240 (B)	0.0140	0.0065	0.0109	0.0048
(10)	mdb241 (M)	0.0075	0.0050	0.0110	0.0019

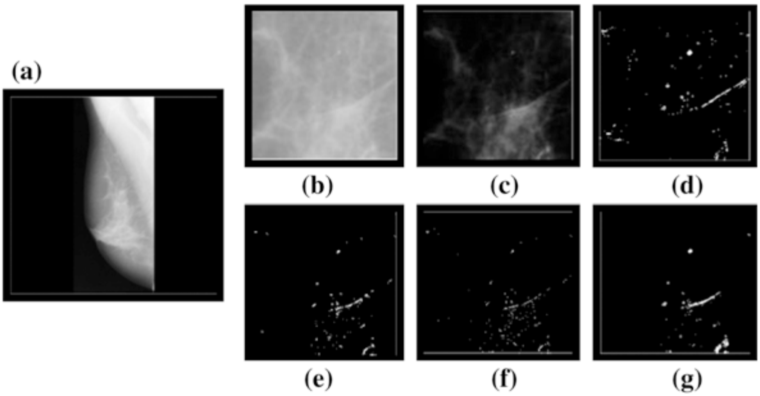


Fig. 3 **a** Original mammogram (mdb213) with calcifications embedded in a background of fatty-glandular tissues. **b** Extracted ROI. **c** Enhanced ROI. Edge maps obtained using: (**d**) ED-I, (**e**) ED-II, (**f**) ED-III, (**g**) ED-IV

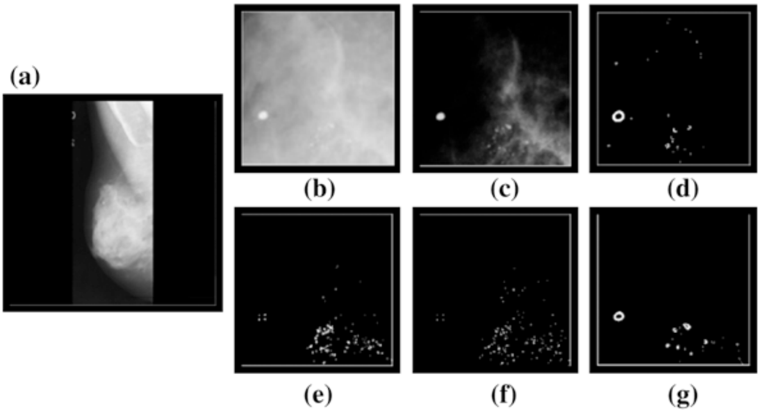


Fig. 4 **a** Original Mammogram (mdb223) containing micoclacification clusters in a dense backgorund. **b** Extracted ROI. **c** Enhanced ROI. Edge maps obtained using: (**d**) ED-I, (**e**) ED-II, (**f**) ED-III, (**g**) ED-IV

which is suitable to trap calcifications of size 25–30 pixels. On the other hand, the second pass of this edge detector can detect the tiny scattered calcifications using a structuring element (of size 3 × 3). Therefore, both clustered as well as aligned calcifications are recovered in edge map of Fig. 2g and are well quantified by the proposed evaluation measure, in the form of extremely low values of error indices.

Figure 4 shows edge maps for mdb223, which is a mammogram with dense background and also classified as one of the radiology difficult cases. From Fig. 4d–f, it can be analyzed that edge detectors, ED-I to ED-III do not give satisfactory performance for these cases. Yet, ED-IV yield reasonably good and precise results with the least values of *MSE* and *MAE* upon reconstruction estimation. From the simulation results obtained and their corresponding evaluations, it can be ascertained that proposed reconstruction estimation function has proved to be a suitable evaluation method for edge maps, as it can return precise data to compare the performance of edge detectors.

With the data obtained in Tables 1 and 2, it can be inferred that the magnitude of the errors indices (*MSE* and *MAE*) are large for edge maps obtained using ED-I, ED-II and ED-III, but it tend to decrease with the application of ED-IV. Also, ED-II can be rated to produce performance nearly comparable performance to ED-IV in case of mammograms with fatty and fatty-glandular breast tissues, but is the performance degrades for dense mammograms. The proposed evaluation scheme quantifies that the performance of ED-IV is significantly improved in comparison to other gradient based morphological edge detectors (ED-I to III) proposed for detection of calcifications.

5 Conclusion

Identification of discontinuities in an image during the reconstruction process becomes the primary essence of edge detection algorithms. Owing to the functional restriction of PFOM in evaluating real images; the compulsion of a non-reference assessment procedure for edge maps of mammograms is highlighted in this work. The advantage of the proposed evaluation method lies in the fact that the interpolation procedure used in performing reconstruction estimation helps in precise and effective generation of gray-scale enhanced images from the binary edge-maps along with retention of image sharpness. Therefore, a clear demarcation is observed in the performance of edge detectors on mammograms with different category of background tissues. The subjective analysis of mammogram edge-maps yields 85 % coherence with the proposed performance measure. Hence, the new version of reconstruction based evaluation will serve as an effective tool to benchmark the performance of various edge enhancement algorithms for detection of calcifications. This will further aid in extraction of features in mammographic images, improving diagnostic detection of breast cancer.

References

1. Rovere GQ, Warren R, Benson JR (2006) Early breast cancer from screening to multidisciplinary management, 2nd edn. Taylor and Francis Group, Florida
2. Canny JA (1986) Computational approach to edge detection. IEEE Trans Pattern Anal Mach Intell 6:679–698 (PAMI-8)
3. Woodhall M, Linquist C (1997) New edge detection algorithms based on adaptive estimation filters. In: 31st asilomar IEEE conference on signals systems & computers, vol 2. pp 1695–1699
4. Sharifi M., Fathy M, Mahmoudi MT (2002) A classified and comparative study of edge detection algorithms. In: Proceedings of the international conference on information technology: coding and computing (ITCC.02), (2002)
5. Strickland RN, Hahn H (1996) Wavelet transform for detecting microcalcifications in mammograms. IEEE Trans Med Imaging 15(2):218–229
6. Dominguez JQ, Cortina-Januchs MG, Jevti'c A, Andina D, Barron-Adame JM, Vega-Corona A (2009) Combination of nonlinear filters and ANN for detection of microcalcifications in digitized mammography. In: Proceedings of the international conference on systems, man, and cybernetics, USA, pp 1516–1520
7. Papadopoulos A, Fotiadis DI, Costaridou L (2008) Improvement of microcalcification cluster detection in mammography utilizing image enhancement techniques. Comput Biol Med 38(10):1045–1055
8. Wu Z, Yuan J, LV B, Zheng X (2010) Digital mammography image enhancement using improved unsharp masking approach. In: Proceedings of 3rd international congress on image and signal processing (CISP), vol 2. pp 668–672
9. Polosel A, Ramponi G, Mathews VJ (2000) Image enhancement via adaptive unsharp masking. IEEE Trans Image Process 9:505–510
10. Yang YB, Shang HB, Jia GC, Huang LQ (2003) Adaptive unsharp masking method based on region segmentation. Optics Precis Eng 11:188–191

11. Stojic T, Reljin I, Reljin B (2005) Local contrast enhancement in digital mammography by using mathematical morphology. In: International symposium on signals, circuits and systems (ISSCS), vol 2. pp 609–612
12. Stojic T, Reljin B (2010) Enhancement of microcalcifications in digitized mammograms: multifractal and mathematical morphology approach. FME Trans 38(1):1–9
13. Dominguez JQ, Sanchez-Garcia M, Gozalez-Romo M, Vega-Corona A, Andina D (2009) Feature extraction using coordinate logic filters and artificial neural networks. In: 7th international conference on industrial informatics (INDIN'09), pp 645–649
14. Santhaiah Ch, Babu GA, Rani MU (2009) Gray-level morphological operations for image segmentation and tracking edges on medical applications. Int J Comput Sci Network Secur 9(7):131–136
15. Bhateja V, Devi S (2011) A novel framework for edge detection of microcalcifications using a non-linear enhancement operator and morphological filter. In: Proceedings of 3rd international conference on electronics & computer technology (ICECT-2011), Kanyakumari (India), vol 5. pp 419–424
16. Shin M, Goldgof D, Bowyer K, Nikiforou S (2001) Comparison of edge detection algorithms using a structure from motion task. IEEE Trans Syst Man Cybern B 31(4):589–601
17. Yitzhaky Y, Peli E (2003) A method for objective edge detection evaluation and detector parameter selection. IEEE Trans Pattern Anal Mach Intell 25(8):1027–1033
18. 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
19. Trahanias P, Venetsanopoulos A (1996) Vector order statistics operators as color edge detectors. IEEE Trans Syst Man Cybern B 26(1):135–143
20. Pratt W (1991) Digital image processing, 2nd edn. Wiley, New York
21. Yitzhaky Y, Peli E (2003) A method for objective edge detection evaluation and detector parameter selection. IEEE Trans Pattern Anal Mach Intell 25(8):1027–1033
22. Devi S, Soni A, Thakur N, Das M (2010) Image quality metric for coupled noise and distortion. In: 4th international conference on advanced computing and communication technologies (ICACCT-2010), India
23. Carlsson S (1988) Sketch based coding of gray level images. Signal Process 15(1):57–83
24. Elder J (1999) Are edges incomplete? Int J Comp Vis 34(2/3):97–122
25. Govindarajan B, Panetta K, Agaian S (2008) Image reconstruction for quality assesment of edge detectors. In: Proceedings of IEEE international conference on systems, man, and cybernetics, pp 691–696
26. Govindarajan B, Panetta K, Agaian S (2009) A non-reference measure for objective edge map evaluation. In: Proceedings of IEEE international conference on systems, man, and cybernetics, pp 4563–4568
27. Carlsson S (1988) Sketch based coding of gray level images. Signal Process 15(1):57–83
28. Suckling J, et al. (1994) The mammographic image analysis society mammogram database. In: Proceedings of 2nd international workshop digital mammography,U.K, pp 375–378

Automatic Detection of Zebra Crossing Violation

Jamini Sampathkumar and Kumar Rajamani

Abstract This paper attempts to provide a solution to detect a particular traffic violation rule—that of stopping on a zebra crossing at a traffic signal, instead of behind it. This violation of norm has previously neither been tackled nor attempted to be resolved. The solution involves a multi layer pipelined stages of pre-processing that include binarizing the image, filtering (that is median and Laplacian filtering), and morphological processing which help to identify if a vehicle that stopped at a traffic signal is halted on the zebra crossing or before it. A template of the zebra crossing that is under investigation is stored in the system for analysis. This template is devoid of any vehicular traffic and captures only the markings of zebra crossing. At runtime, the image captured is sent to a detector which combines the analysis of correlation through three different correlation metrics and subsequently detects any violation. This approach has been tested on manually acquired 50 test data spread over two different zebra crossings on Indian roads with an accuracy of over 90 %. This approach is simple, yet efficient and robust.

Keywords Traffic violation detection · Correlation · Mean square distance · Morphological operations · Segmentation · Zebra crossing · Computer vision

J. Sampathkumar (✉)
SAP Labs India, Bangalore, India
e-mail: jamini.sampathkumar@sap.com

K. Rajamani
Sathya Sai Institute of Higher Medical Sciences, Bangalore, India
e-mail: kumartr@gmail.com

Mohan S. and S. S. Kumar (eds.), *Proceedings of the Fourth International Conference on Signal and Image Processing 2012 (ICSIP 2012)*, Lecture Notes in Electrical Engineering 222,
DOI: 10.1007/978-81-322-1000-9_47,

1 Introduction

The growth of an economy brings with it a rapid rise in vehicle density in the cities. Statistics reveal that the number of new vehicles plying the Indian roads have risen exponentially [1]. This results in a corresponding increase in traffic-related problems and the difficulty in handling traffic violations. Law enforcement authorities are looking at Information Technology (IT) as a way to combat this issue.

As a result, this has motivated research in the area of effectively employing IT as a methodology to address management of vehicle regulation. Significant effort has been directed in identifying and automating the detection of various traffic violations.

A lesser known, and investigated offense at a traffic signal, is the stopping of a vehicle on the markings of zebra crossing. Rules demand that a vehicle at a traffic signal stop before the zebra crossing to enable pedestrians to use the marked area to cross the road. However, investigations by the traffic regulatory authorities in India show that this rule is often not obeyed.

A traditional approach would involve a traffic official to manually monitor the traffic at the zebra crossing to identify any violation in order to report it. This involves deploying of man power, besides being tedious and cumbersome. Today, using information technology, this task can be automated using sophisticated algorithms.

Our investigations reveal that research in this domain has been undertaken, especially to assist the visually impaired to detect and navigate the zebra crossing. Ludwig et al. [2] used Adaboost approach to detect presence of zebra crossings.

Stephen Se [3] used pose and vanishing lines, along with an existing application to assist visually impaired. Uddin et al. [4] used projective invariants for the zebra crossing detection process. There has also been some interest in traffic surveillance [5, 6] at zebra crossings. However, there has been no explicit work in the area of detection of traffic violation as a result of stopping of vehicles on zebra crossings.

This paper presents a simple approach for automatic detection of this traffic violation. The main objective is the detection of vehicles “stopping on zebra crossings”. The current algorithm has been developed specifically for use on the Indian roads. The input to this algorithm is an image of the zebra crossing at the traffic signal. The approach describes the pre-processing required in order to eliminate noise in the input image, followed by a comparison of the input image to be tested for violation with a template image. The algorithm works in an offline mode. The following section presents the methods used in this approach. Further, the results and the conclusions are elucidated. Finally, the extensions on this algorithm are elaborated.

2 Method

The following section describes the methodology employed in order to detect a violation at a zebra crossing. The process is composed of two stages—Pre-processing and Detection. The first stage contains a series of closely-orchestrated steps that eliminates noise in the input image to facilitate violation detection. The output of this stage acts as the input to the second stage where the actual detection of a violation is done.

2.1 Pre-Processing

Following are the steps in the pre-processing of the input image. The input to this stage is the image of the zebra crossing that is captured at a traffic signal when the "Stop" signal is on. This image typically contains both the zebra crossing and the vehicular traffic. The second image is that of the zebra crossing in a "zero-traffic" scenario which serves as the "Control" image. All pre-processing is based on this control image.

2.1.1 Registration

The input image is registered with the control image. Registration is the process of transforming different sets of data into one coordinate system. This is necessary in order to be able to compare or integrate the data obtained from these different measurements. Here, registration is done in order to avoid any error in measurement and detection due to shake of the camera or change in angle of image during acquisition.

The template image is the source image and the input image is the sensed image. Image registration involves spatial transformation of the target image to align with the reference image. There are various approaches to image registration [7–12]. The mechanism adopted in this paper is based on point set registration.

The steps in Point-Set based Image registration are:

1. Detecting and refining input and template images.
2. Clearing of blobs and islands by morphological processing.
3. Matching the resulting sets of points using a point set registration algorithm with the Gaussian mixture models to estimate the correspondence [13].
4. Estimating the correspondences from the transformation.

2.1.2 Binarizing

Binarizing is done in order to facilitate segmentation and comparison of input image with the control, without any reference to color. This makes the algorithm invariant to the color of the binarizing environment (traffic signal containing vehicles of different colors).

The following process is followed to binarize the input image:

1. If the image is a color image then it is converted to grayscale. This is achieved by eliminating any information in the hue and saturation while retaining the luminance.
2. Otsu's method [14] is then used to achieve a global image threshold.

2.1.3 Segmentation

Segmentation is done in order to extract the zebra crossing from the image by a technique of clearing border. This is achieved by suppressing structures that are lighter than their surroundings and connected to the images' border. The connectivity is 8 for a 2-dimensional input-image being segmented. This approach is used as binarizing the image creates a sharp variation in intensity at the boundaries. The connectivity of the applied mask is 8 for the input 2-dimensional image being segmented.

2.1.4 Morphological Operation

Segmentation tends to erode the borders of the image. Hence, morphological operations are used to smoothen and dilate the image to restore the eroded or thinned borders. The following structuring element is used to perform dilation:

$$\begin{bmatrix} 00000 \\ 01110 \\ 01110 \\ 01110 \\ 00000 \end{bmatrix}$$

2.1.5 Laplacian

The image obtained after performing the above steps contains reduced noise. However, the clarity of the image is reduced owning to segmentation using border-clearing technique followed by smoothening. In order to further enhance the image quality in order to facilitate comparison with the existing template library, a Laplacian filter is applied on the image which highlights its edges. The Laplacian filter is computed as follows:

$$\nabla^2 = \frac{4}{\alpha+1}\begin{bmatrix} \frac{\alpha}{4} & \frac{1-\alpha}{4} & \frac{\alpha}{4} \\ \frac{1-\alpha}{4} & -1 & \frac{1-\alpha}{4} \\ \frac{\alpha}{4} & \frac{1-\alpha}{4} & \frac{\alpha}{4} \end{bmatrix}$$

where alpha takes the value 0.2 since Laplacian is calculated in double format.

Upon subtraction of the Laplacian from the original image, we obtain an enhanced image with sharper and more defined edges that aid accurate correlation.

2.2 Detection

Once the image is pre-processed and devoid of any noise or undesirable artifacts, the actual detection of violation is done. The methodology employed in order to detect a violation is a comparison between the control image of the zebra crossing and the pre-processed input image. The multiple methods of comparison between images evaluated are as follows:

2.2.1 Subtraction of Template and Input Image

By subtraction followed by an evaluation of how different the result is different from a "zerointensity" image, the degree of difference between the input image and control image can be defined. However, this technique is vulnerable to even a single pixel shake or movement of input image with template.

2.2.2 XOR of Template and Input Image

The result of the XOR operation between in the input and control image gives a measure of the "inequality" between the images. Although this approach is less vulnerable than subtraction for translation between input image and template, it is still susceptible to movement.

2.2.3 Correlation of Template and Input Image

The coefficient of correlation computed between the input image and the control gives a reliable measure of similarity between the images. However, it is susceptible to illumination variances.

r = corr2(A,B) [15]

The correlation coefficient is calculated as follows:

$$r = \frac{\sum_m \sum_n (A_{mn} - \overline{A})(B_{mn} - \overline{B})}{\sqrt{\left(\sum_m \sum_n \left(A_{mn} - \overline{A}\right)^2\right)\left(\sum_m \sum_n \left(B_{mn} - \overline{B}\right)^2\right)}}$$

where = mean2 (A), and = mean2 (B)

2.2.4 Mean Square Distance Comparison

Yields good result, but is also susceptible to illumination variances. The mean square distance between 2 pixels in 2 images is computed as follows:

M = mean (A) [1]

This returns the mean values of the elements along different dimensions of an array.

$$M = \frac{1}{n} \sum_{i=1}^{n} (x_i)'$$

where n is the number of elements in that dimension of the array.

Once the mean square distance between the reference point and every pixel in the image is computed, a comparison with the threshold is made to evaluate the similarity between the input images (that has already been registered) with the control.

2.2.5 Standard Deviation

Gives a reliable result, but is sometimes vulnerable to appearance of other artifacts.

The standard deviation is a 2-dimensional standard deviation that is measured as follows:

s = std (X) [15]

$$s = \left(\frac{1}{n-1} \sum_{i=1}^{n} (x_i - \overline{x})^2\right)^{\frac{1}{2}}$$

where

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

And n is the number of elements in the sample. The result ‘s’ is the square root of an unbiased estimator of the variance of the population from which ‘x’ is drawn, as long as ‘x’ consists of independent, identically distributed samples.

This paper combines the comparison methods discussed in 3–5 above, to detect a violation in the zebra crossing. The correlation, mean square distance and the standard deviation are computed between the input image and the control image. The decision of a *violation or no-violation* is then made based on individual thresholds set for the measurements thus obtained. These thresholds are set empirically. First, a correlation coefficient below the set threshold is considered as low similarity between the images. Second, a mean square distance lower than the set threshold is considered as low degree of similarity between the images. Third, a Standard deviation value of the input image being lesser or greater than the standard deviation of the control image by the threshold is considered as low similarity. The converse of the comparisons also applies. Finally, the cumulative results of the three methods of comparison are taken into consideration to decide if a violation exists.

2.3 Algorithm

Step 1: Registering the input-image with the template. This ensures that errors in detection that arise from linear or angular shift between the input and control image do not occur.

Step 2: Binarizing the image. If the input image is a color image, it is converted to a black and white image.

Step 3: Using segmentation to clear boundary to extract the Zebra crossing region. Region of interest here is the marked area of the zebra crossing alone.

Step 4: Using Morphological Technique (Dilation) on extract zebra crossing region. This ensures that the segmentation process does not leave and eroded image for detection.

Step 5: Smoothening image to reduce noise. This morphological process further reduces the adverse effects of segmentation due to border clearing.

Step 6: Filtering original image using Laplacian in a double format which results in a negative image.

Step 7: Obtaining enhanced image by subtracting original image from the Laplacian. This enhances the image facilitating more accurate detection.

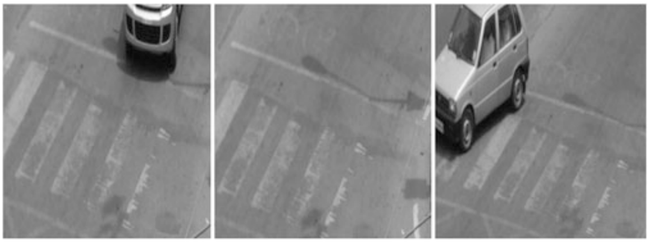


Fig. 1 The variability of the input test data and invariance to rotation or scaling

Step 8: Resizing the Extracted image of zebra crossing is to the size of template. This is necessary to make the image suitable for detection using metrics that are size dependent.

Step 9: Comparing the resized image with the control image using the maximum correlation result to obtain a correlation coefficient and determining its relation to the threshold.

Step 10: Comparing the resized image using the Standard Deviation result followed by determining its relation to the Standard Deviation of the control.

Step 11: Comparing the resized image with the control image using the mean square distance result and determining its relation to the threshold.

Step 12: Determining a violation or no-violation after considering the cumulative of the above three results.

Step 13: Documenting the result for any existence of violation.

Figure 2 depicts the different stages in the implementation of the above algorithm.

3 Results

The algorithm has been tested on manually obtained datasets from the Indian roads.

3.1 Test Data

The images were captured manually from a camera whose resolution and zoom level is known. Here, a camera with a resolution of 12 mega-pixels, zoom of 3.00 × and a maximum aperture of 2.8 was used. The test data consisted of 50 images of Zebra crossings with and without violations. The algorithm is invariant to the color, size, liner and angular shift of the input image. The test data (Fig. 1) captures some of the many variations.

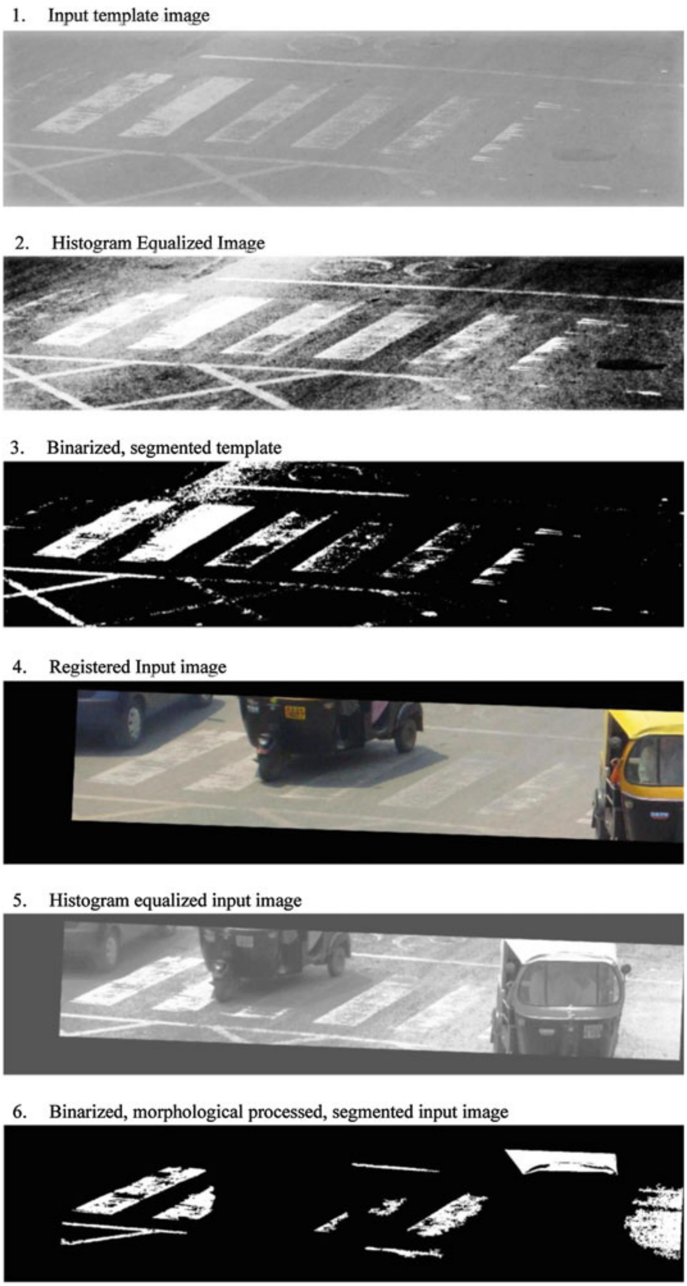


Fig. 2 Depicts the different stages in the flow of the algorithm and the respective output. In this scenario, the combined metric detects a violation. **a** Input template image. **b** Histogram equalized image. **c** Binarized, segmented template. **d** Registered input image. **e** Histogram equalized input image. **f** Binarized, morphological processed, segmented input image

3.2 Variability of Test Data

Figure 1 demonstrates the variability of the input data that was observed in the 50 datasets that were acquired manually.

The can be scaled or rotated as compared to the control image.

3.3 Implementation of Algorithm

This section presents the results at the various stages in the algorithm computation of the algorithm.

3.4 Results for Tested Data

Results for the various methods were tabulated, based on which a cumulative result was taken into consideration. The algorithm was tested on all the manually acquired 50 images of zebra crossings irrespective of violation. The accuracy in recognition for the algorithm was 90 %. The accuracy levels can be improved further by using a more elaborate set of templates during the detection. The test scenario on Indian roads posed a wide variation in environs. Moderately varying illumination, variation in traffic and faded Zebra crossings were the challenges faced.

4 Conclusion

This paper presents an innovative and elegant approach for detection of violation at zebra crossing. The advantages of this technique are its sophisticated, streamlined preprocessing approach, ease in implementation and realization of the algorithm and a highly robust and rapid detection of any violation. It is ideally suited for Indian environs. The extension of this technique is presently being carried out to also include the recognition of the license plate of the vehicle that is violating the traffic regulation in order to facilitate punitive action against the offender.

References

1. SP Gupta, India Vision 2020. http://www.Planniningcommission.nic.in/reports/genrep/pl_vsn2020.pdf
2. Lausser L, Schwenker F, Palm G (2008) Detecting zebra crossings utilizing AdaBoost. In: ESANN 2008, Belgium
3. Stephen S (2000) Zebra-crossing detection for the partially sighted. In: CVPR 2000
4. Uddin MS, Shioyama T (2005) Detection of pedestrian crossing using bipolarity and projective invariant. In: MVA2005, Japan
5. Eikvil L, Huseby RB (2001) Traffic surveillance in real-time using hidden Markov models. In: Proceedings of the 12th Scandinavian conference on image analysis, Bergen
6. Setchell CJ (1997) Applications of computer vision to road traffic monitoring. Ph.D thesis, Computer Vision Group, Bristol University
7. Michalewicz Z (1996) Genetic Algorithms + Data Structures = Evolution Programs. 3rd edn. Springer-Verlag, Berlin, Heidelberg
8. Goshtasby A (2005) 2-D and 3-D image registration for medical, remote sensing, and industrial applications. Wiley Press, New Hoboken
9. Simonson K, Drescher S, Tanner F (2007) A statistics based approach to binary image registration with uncertainty analysis. IEEE Pattern Anal Mach Intell 29(1):112–125
10. Domokos C, Kato Z, Francos J (2008) Parametric estimation of affine deformations of binary images. In: Proceedings of IEEE international conference on acoustics, speech, and signal processing, 2008
11. Goshtasby A (1986) Piecewise linear mapping functions for image registration. Pattern Recogn 19:459–466
12. Goshtasby A (1988) Image registration by local approximation methods. Image Vis Comput 6:255–261
13. Jian B, Vemuri BC (2011) Robust point set registration using gaussian mixture models. In: IEEE PAMI, 2011
14. Otsu N (1979) A threshold selection method from gray-level histograms. IEEE Trans Syst Man Cybern 9(1):62–66
15. Matlab Central. http://www.mathworks.in/matlabcentral/index.html
16. Brown LG (1992) A survey of image registration techniques (abstract). ACM Comput Surv (CSUR) Arch 24(4):325–376
17. Auty GW, Corke PI, Dunn PA, MacIntyre IB, Mills DC, Simons BF, Jensen MJ, Knight RL, Pierce DS, Balakumar P (1998) Vehicle Monitoring System. Patent US5809161
18. Sedgewick R (1998) Algorithms in C, 3rd edn. Addison-Wesley, pp 11–20

Fabric Paint Environment Using 3D Haptics for Vocational Training

Shreyas Valmiki and Kumar Rajamani

Abstract Fabric art is an important aspect of fashion designing. Though there is lot of talent in this field, only a few are professionally trained and take it up as a vocation due to financial or physical disabilities. This paper provides a solution for such aspiring individuals to train in this field. Fabric painting is the art of painting on different textures and contours, hence making it a tough occupation because of the variety of materials that are available to be worked on; this is where the haptics technology comes to use. A three-dimensional haptic device, such as the Novint Falcon, converts hand gestures to virtual 3D motion. It is also equipped with force feedback technology which allows the user to "feel" the virtual object in contact within a cursor in a virtual environment. To create good applications using this technology a comprehensive toolkit is needed that efficiently puts to use the various features of this device to create realistic tutoring applications. CHAI 3D is one such toolkit and with the ability to provide support to multiple haptic devices and with the lightweight rendering prowess of OpenGL technology, it is the choice made for this purpose. CHAI 3D bridges two components—the 3D graphics and the haptic technology to simulate interaction with real-world objects. This paper provides an elegant approach to painting on a three-dimensional canvas, as opposed to the usual 2D canvas, that can translate and rotate around any axis. To add to that, the paper proposes to use the force feedback technology that is built into most 3D haptic devices to allow the user to "feel" every stroke of the brush on the virtual piece of clothing that can vary with texture and contour making the learning experience of an aspiring fabric artist a more realistic one.

S. Valmiki (✉)
Department of Computer Science, Polytechnic Institute of New York University, New York, USA
e-mail: shreyasvalmiki@gmail.com

K. Rajamani
Sathya Sai Institute of Higher Medical Sciences, Bangalore, India
e-mail: kumartr@gmail.com

Mohan S. and S. S. Kumar (eds.), *Proceedings of the Fourth International Conference on Signal and Image Processing 2012 (ICSIP 2012)*, Lecture Notes in Electrical Engineering 222, DOI: 10.1007/978-81-322-1000-9_48, © Springer India 2013

Keywords Fabric art education · Novint falcon · CHAI 3D · 3D graphics and haptic technology · OpenGL · Three-dimensional canvas · Force feedback mechanism · Texture and contour

1 Introduction

There is no dearth of talent in India and one such talent is fabric painting. This is the art of painting on fabrics of different textures and contours. Painting on fabric also requires a good hand at cutting the fabric to suit the demand which requires drawing prowess as well. Art, especially fabric art, is a very important aspect of fashion designing [3]. People in the cities have access to supplies and also get good training from institutes to take up fabric painting as a profession, but that is not the case with the people in remote areas of India, people who want to learn but cannot because of financial or physical disabilities. This paper brings in the possibility of providing a more realistic and efficient art education to the masses who aspire to take up fabric painting or garment designing as their profession.

There are several input devices we see today in the market, one of the most common being the simple computer mouse. The mouse is, even today, used effectively for vocational applications such as fabric painting. Other training equipments for painting include the tablet that converts gestures made on the pad to patterns on the screen. Both the formerly mentioned devices work in two-dimensions, although effective for training people who aspire to be canvas artists, does not provide an efficient medium to learn painting on contoured and textured surfaces. This can be achieved by a three-dimensional haptic device that also provides force feedback to the user. A 3D haptic device is capable of converting hand gestures to three-dimensional motion in a virtual environment and the capability of force feedback in such devices mimic the sense of touch. The aforementioned features of a three-dimensional haptic device have already been applied in various fields such as medicine [5] and gaming [6]. There are a few notable manufactures of such devices: one such being SensAble Technologies [9]. PHANToM [10] is their 3D haptic device with force feedback mechanism built in. Novint [12], another manufacturer of 3D haptic devices entered the haptic market by introducing it in the gaming industry. The Novint Falcon was widely accepted for gaming as it was made at an affordable price and catered to a wider audience [4]. Apart from being the more affordable device, it is considered superior to the PHANToM [7]; one such feature being, it features higher sensitivity with respect to force feedback than the PHANToM; the Novint Falcon was chosen the very reason—for making the fabric painting experience more realistic, higher sensitivity to force applied and the feedback to the user is important.

There are many tools and libraries which could be used to create applications for these devices, such as: H3D, HAPI, CHAI 3D, Novint HDAL SDK, etc. Of these, H3D, HAPI and CHAI 3D are open-source. The Novint HDAL SDK though, is a toolkit for creating good start-up applications for Novint Falcon. Apart from being device specific, it is not open-source. Extensive software applications cannot be created as it is not as flexible as others in the competition. H3D [8] is open source API and use good graphics rendering technologies such as OpenGL and X3D but are specific to three-dimensional haptic devices only. CHAI 3D was developed by Francois Conti, Federia Barbagli et al. at Stanford University [10]. It is actively being enhanced with lots of research and development. It is a comprehensive toolkit for development of applications for haptic devices. It uses the lightweight rendering ability of OpenGL and is not specific to three-dimensional haptic devices; applications for 2D haptic devices can also be created. It also provides support for multiple haptic devices at the same time which is a big advantage. CHAI 3D provides a virtual haptic simulator which can be used to simulate the movement made by a haptic device. Hence applications can be created even without the presence of the actual device. This advantage made it an ideal choice for our experiments.

The realm of Computer Graphics is still a theme under development in India. While countries like Japan are making great strides in the expansion of this field, graphic studios have only recently sprung about in India. While 3D graphics proves to be a fairly routine modus operandi in the entertainment industry, it is the Haptics technology that is taking it to the next level. The experience of “touch” using the force feedback technology makes this field of research very enthralling. While most applications in the market provide a two-dimensional platform to paint on, it is the objective of this paper is to offer the exotic idea of a three-dimensional canvas that can rotate and translate around any axis. The force feedback mechanism of the most 3D haptic devices allow the user to “feel” every stroke of the brush on the canvas, and the “canvas” can vary from a flat piece of cloth to any complex piece of clothing. In this paper, we present the first experiment results on the feasibility of 3D haptic enabled fabric painting environment. This application when developed completely would highly assist in teaching fabric painting to people in rural areas. It would also help the rural people to practice and gain experience. Once they get sufficient experience they can seek a profession to be gainful employed.

2 Method

In this section the techniques that were used in building the 3D haptic enabled fabric painting environment is elaborated. First the concept of scenegraph and haptic thread is explained. This is followed by the mechanism by which the virtual scene is composed and rendered. Finally the aspects of collision detection and details of getting the graphics drawn on the canvas are detailed.

2.1 The 3D Scenegraph

A 3D scenegraph consists of the world. World, in terms of CHAI 3D, is a virtual environment that is generated by OpenGL that acts as a container for all the objects on the three-dimensional scene. A "world" is this virtual environment that houses all the 2D widgets and the 3D meshes. These objects take a finite volume and position in the world. The world in this case extends to infinity in all the coordinates. A camera in reality is a device that captures and records the reflected light bounced off the objects in front of it. The function of a camera in the virtual world is exactly the same. It captures and displays the images that are formed by the light bouncing off the objects in the world. The light source in the virtual world illuminates the objects. This is the light that makes the scenes visible on the screen. The camera captures the reflected light in the world and a user can only view the part of the world that is focused by the camera. It is initialized by setting its view position, the target or the look-at position and then camera should know which coordinate forms the up vector. CHAI 3D also allows a set of clipping planes for the camera. The camera is programmed to show only the objects that are between the 2 clipping planes. The camera again is made the child of the world.

CHAI 3D allows multiple sources of light present. It also allows different coloured lights for illumination of the world. It is usually declared as the child of the camera so that it illuminates the objects in the world that the camera is pointed at. It also moves with the camera, if the camera is made to move from one position to another. Figure 1 shows the activity diagram of the 3D scenegraph.

2.2 The Haptics Thread

The haptic device, though tangible, is generally represented as a virtual dot in a 3D scenegraph. Since CHAI 3D allows multiple haptic devices to participate at the same time, each device is represented by a handler for identification. The haptic drivers and CHAI 3D translates the physical movement of the device into to the virtual movement of the handler. The measurements and distances in the virtual scene generally do not match the actual movement of the device. Hence a conversion factor is applied to give a realistic experience.

A fresh thread for the device alone is necessary as every movement of the haptic device is to be portrayed on the scenegraph seamlessly. A mathematical function called a Jacobian is then used to determine the position of a three-dimensional cursor in Cartesian coordinates based on the positions of the arms. The Jacobian matrix is represented as follows:

Given a set $y = f(x)$ of n equations in n variables $x_1, \ldots\ldots, x_n$, written as

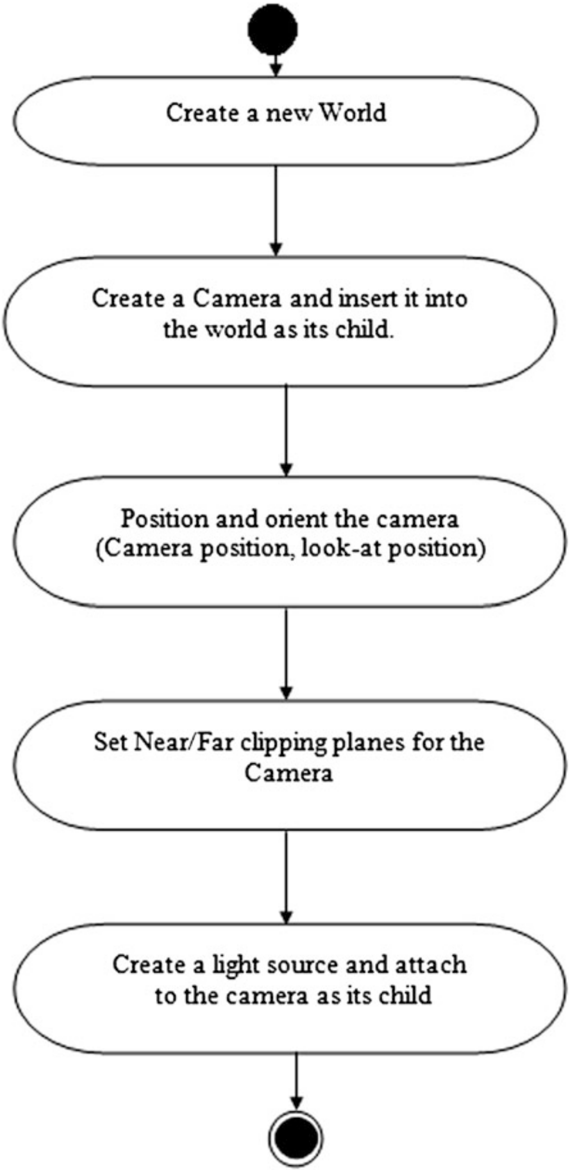


Fig. 1 3D scenegraph activity diagram

$$y = \begin{pmatrix} f_1(x) \\ f_2(x) \\ \vdots \\ f_n(x) \end{pmatrix}$$

or explicitly as

$$\begin{cases} y_1 = f_1(x_1, \ldots\ldots, x_n) \\ \vdots \\ y_n = f_n(x_1, \ldots\ldots, x_n) \end{cases}$$

the Jacobian matrix, sometimes simply called "the Jacobian" [11] is defined by

$$J(x_1, \ldots, x_n) = \begin{pmatrix} \frac{\delta y_1}{\delta x_1} & \cdots & \frac{\delta y_1}{\delta x_n} \\ \vdots & \ddots & \vdots \\ \frac{\delta y_n}{\delta x_1} & \cdots & \frac{\delta y_n}{\delta x_n} \end{pmatrix}$$

The determinant of J is the Jacobian determinant and is denoted

$$J = \left| \begin{matrix} \delta\ (y_1, \ldots, y_n) \\ \delta\ (x_1, \ldots, x_n) \end{matrix} \right|$$

The movement of the device is captured through an action listener and the screen is updated to show its current position. The scenegraph with the 3D and the 2D objects constitute the main thread, so that the interaction between the device cursor and the objects can be shown on the screen as and when it happens. When the cursor comes in contact with an object (with an applied texture and density) the forces between them are captured using the 3 servo motors. These values are read using the cGenericHapticDevice class. The screenshot in Fig. 2 shows the haptic cursor with a ball shaped graphic called the tool. Also shown on the screen is the reading for the position of the cursor in 3D space and the force that is being applied in each of the coordinates.

A tool is the virtual representation of the haptic device on the screen. This tool is an object of the cGeneric3dofPointer class. The "tool" is generically used to refer to any of the haptic device, including the Falcon or the virtual simulator. The default tool in CHAI 3D is represented by a small sphere this is also called the proxy. When a force is applied on to the Falcon by the application, the tool separates into 2 spheres with a line in between. The length of the line is proportional to the force applied. The radius of the sphere can be set here. The tool is, by default, masked by an orange sphere, which in this case, is to be disabled.

One of the important attributes of the tool that needs to be set is the workspace radius. The workspace radius determines how the physical workspace of the haptic device is mapped to the virtual workspace on the screen. In this case, the virtual

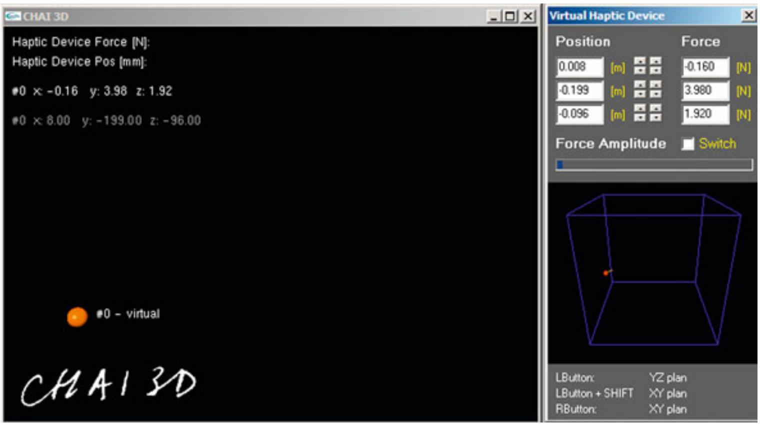


Fig. 2 Hepatic cursor with virtual simulator

workspace is made equal to the physical workspace hence the measurements will be the same both in the physical and the virtual environments.

The haptic device consists of a switch that can be used to our advantage. When the switch is pressed, the tool is programmed to grab the canvas instead of the pen. When the grip is moved when the switch is on, the canvas can be rotated using the pen as the axis. The canvas can also be translated when the pen is sufficiently far away from the pen. This is done by changing the rotation of the canvas.

When the switch is on, the pen is disabled and the control is only on the canvas. When the switch is off, the Falcon grabs the pen. All these functionalities of the falcon are enabled in the haptics thread and are necessary to carry out the various tasks in the fabric painting tutor.

2.3 Virtual Scene Composition

The virtual scene consists of the 3D objects that can be seen on the screen. In the paint application, there are 2 objects: the pen and the canvas. Both the canvas and the pen are of '3ds' format. The 2 objects are loaded dynamically each time the application starts to save memory. Any mesh of the '3ds' format is stored in an object of the class cMesh which is an internal CHAI 3D class specially built for manipulating 3D meshes. 3D objects that have to be displayed on the screen is made a child of the "world". The world, when rendered, renders all its children. The position of the mesh and the initial rotation is defined. The canvas is scaled to fit the screen and the boundary box for the mesh has to be computed to calculate collision and to assign material properties. The virtual scene also consists of the material properties of 3D objects.

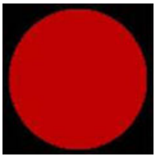


Fig. 3 Ambient property

Fig. 4 Diffuse property

Fig. 5 Specular property

2.4 Material Properties

The stiffness factor is one among the most important material property. It is used to set the stiffness of the canvas, in this case, the fabric. This property is responsible for asserting a certain amount of force when the tool comes in contact with the canvas.

There are totally 3 types of material properties: ambient, diffuse and specular. These 3 properties are relative to the light source assigned to the world.

Ambient: Figure 3 shows a 3D sphere lit by ambient light only; it appears to look 2D. Ambient light is the average volume of light that is created by emission of light from all of the light sources surrounding (or located inside of) the lit area.

Diffuse: Figure 4 shows a sphere lit by diffuse light. Diffuse light represents a directional light cast by a light source. Diffuse light can be described as the light that has a position in space and comes from a single direction.

Specular: Figure 5 shows a sphere lit with specular light. Just like Diffuse light, Specular light is a directional type of light. It comes from one particular direction. The difference between the two is that specular light reflects off the surface in a sharp and uniform way

The Specular property is applied to the pen. This property gives the pen a more realistic appearance. The canvas on the other hand is applied with Diffuse property. This is to not produce shine.

2.5 Rendering the Virtual Scene with GLUT

GLUT is a library that has classes and functions that renders the basic user interfaces like windows, menu boxes, check boxes, etc. To initialize the window, the size, the position of the window and initial display mode is defined.

GLUT has various inbuilt functions that take function pointers as arguments and runs in an infinite loop till stopped manually. These functions are mostly action listeners such as key stroke listeners, mouse click listeners or resize window listeners. A function such as updateDisplay is used to make updates to the display as and when the user makes changes in position of the pen or the canvas.

2.6 Collision Detection

The position and the forces are represented in the form of 3D vectors. CHAI 3D has a class named cVector3d that is used for all the vector variables. It contains x, y and z attributes. The force applied and the position of the tool can be determined obtained using getPosition() and getForce() functions of the cGenericHapticDevice class. Collision can be considered to have happened when the force applied on the tool, in this case the pen, is greater than zero. This happens when the pen touches the canvas. Collision is recorded even if the force between the two objects is minute.

2.7 Drawing Lines

A line for a 3D painting application should have 3D properties. CHAI 3D's native function, cShapeLine, does not allow 3D lines; it only renders 2D lines. The customized cShapeLine3d which renders 3D lines is used for the purpose of this paint tool.

Figure 6 represents the activity diagram for drawing lines on the screen.

2.8 Results

The results of the framework that was put together of the 3D haptic enabled fabric painting environment is presented in this section.

The complete workflow is captured in Fig. 6, which shows the activity diagram. This flowchart captures the steps in which the painter makes a stroke on the canvas. The first contact point is recorded as the starting point of the line. The last contact point is recorded as the end point and a line is drawn connecting the two

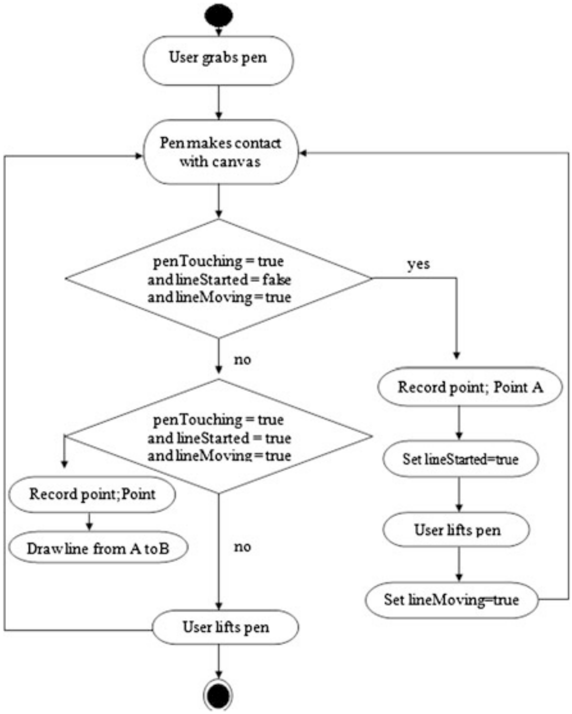


Fig. 6 *Drawing lines* activity diagram

Fig. 7 Rendered 3D pen

points. The line is rendered as part of the 3D environment. Any changes made to the canvas, such as rotation/translation also results in the line being moved in tandem with the canvas.

The haptic device is generally portrayed as a dot. To give the realistic painting experience in a fabric painting application, a pen or stylus is generally needed to be displayed to give a better feel of real-time. Figure 7 shows a fully rendered pen object that could be moved in three-dimensions with the haptic device.

Figure 8 shows the screenshot of the application when the switch is held. When the switch is held, the user can grab the canvas and rotate it or translate it. This is to view a particular piece of drawing in every angle.

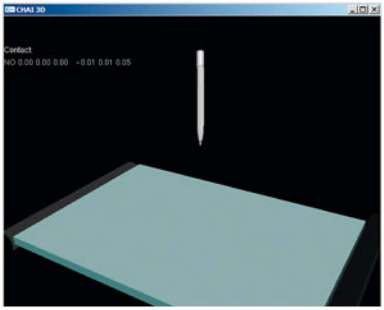


Fig. 8 Switch is on

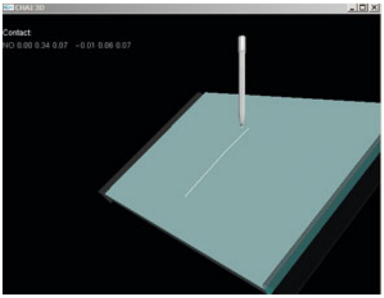


Fig. 9 *Line* drawn on canvas

Figure 9 shows the screenshot of the line being drawn on the canvas. The canvas here is positioned according to the user's liking.

3 Conclusion

A first prototype of 3D haptic enabled fabric painting environment has been presented in this paper. This application can be used by users new to fabric painting with little bit of training. The application has been built taking into account all the real time actions of the painting. This gives the user a very realistic feel to drawing and painting.

The application takes into account all the 3D features of painting such as pressure, force applied, and collision detection. The paper presented here is the first step towards a 3D haptic enabled vocational training material for fabric painting. The enhanced version or prototype is under progress which includes inclusion of various brushes and different colors and textures.

References

1. Baxter W, Scheib V, Lin M, Manocha D (2001) DAB: interactive haptic painting with 3d virtual brushes. In: Proceedings of ACM SIGGRAPH 01, Aug 2001
2. Liang S, Chan ECL, Baciu G, Li RH (2010) Cognitive garment design interface using user behavior tree model. In: IEEE international conference on cognitive informatics (ICCI), Beijing
3. Long X, Li W, Weiyan L (2009) Design and application of fractal pattern art in the fashion design. In: International workshop on chaos-fractals theories and applications (IWCFTA'09), Shenyang
4. Novint Falcon http://www.novint.com/index.php/novintfalcon
5. Kwon DS, Kyung KU, Kwon SM, Ra JB, Park HW, Kang HS, Zeng J, Cleary KR (2001) Realistic force reflection in a spine biopsy simulator, In: Proceedings of the IEEE international conference on robotics and automation (ICRA), Korea
6. Chang D (2002) Haptics: gaming's new sensation, Computer
7. Vanacken L, Boeck J De, Coninx K (2010) The phantom versus the falcon: force feedback magnitude effects on user's performance during target acquisition, HAID 2010 Berlin H3D API, http://www.h3dapi.org/
8. SensAble Technologies, http://sensable.com/PHANToM Haptic device http://sensable.com/products-haptic-devices.htm
9. People at CHAI 3D, http://chai3d.org/people.html
10. Simon CP, Blume L (1994) Mathematics for economists, Norton
11. Gupta N, Yogeshwara K, Vinay Kumar RN, Bisht RS, Ginnela Y, Rajamani K, Achuthan K, Bijlani B (2009) Virtual fabric paint environment using haptics for vocational training, IEEE HAVE 2009, Italy
12. Novint http://www.novint.com/

An Edge Detection Approach for Images Contaminated with Gaussian and Impulse Noises

Ankush Gupta, Ayush Ganguly and Vikrant Bhateja

Abstract Color edge detection is preferred to grayscale edge detection because edges existing at the boundary separating regions of different textures which cannot be detected (in case of grayscale images) if there are no intensity changes. This paper proposes an approach for performance improvement of Hilbert transform based edge detector making it capable of color edge detection in noisy environment. Combining Bilateral filtering with Hilbert Transform produces good results in case of images contaminated with Gaussian and impulse noises. The initiation of the proposed edge detection approach is marked by the transformation of the input color image using RGB color triangle. Computer simulations are performed on noise free images as well as those corrupted with a mixture of Gaussian and impulse noises. Simulation results along with their quality assessment illustrate the effectiveness of the proposed approach in noise free as well as noisy environment.

Keywords Bilateral filtering · Gaussian noise · Hilbert transform · Reconstruction estimation function · RGB color triangle

A. Gupta (✉) · A. Ganguly · V. Bhateja
Departmet of Electronics and Communication Engineering,
Shri Ramswaroop Memorial Group of Professional Colleges,
Faizabad Road, Lucknow 227105, Uttar Pradesh, India
e-mail: ankushh.guptaa@gmail.com

A. Ganguly
e-mail: ayushnascarp@gmail.com

V. Bhateja
e-mail: bhateja.vikrant@gmail.com

Mohan S. and S. S. Kumar (eds.), *Proceedings of the Fourth International Conference on Signal and Image Processing 2012 (ICSIP 2012)*, Lecture Notes in Electrical Engineering 222,
DOI: 10.1007/978-81-322-1000-9_49,

1 Introduction

Image processing involves modifying the attributes of an image which includes performing primitive operations to reduce noise, enhance contrast and sharpen the edges of an image. Edges reflect the local feature differences of images thereby assisting in segregating different objects in the computer vision applications. Edge detection is concerned with locating the significant variations of the gray level in the image and identification of the physical phenomena that originated them. Edge detection, thus becomes an essential pre-processing step for operations as it considerably reduces the amount of information for processing [1]. Color edge detection is preferred to grayscale edge detection because edges existing at the boundary separating regions of different textures which cannot be detected (in case of grayscale images) if there are no intensity changes. In addition, some form of color edge detection is necessarily required to resolve the points that have not been considered while using grayscale edge detection [2]. Gaussian and impulse noises are some of the common noise sources which may be introduced in an image during its acquisition and transmission respectively. These noises not only trammel human interpretation but also cause complexity in computational operations. The noise corrupted images generally suffer from low signal-to-noise ratio and may not be suitable for further processing without removing or reducing the noise content. Traditional edge detection operators like Sobel and Roberts [3] were based on the concept of 'enhancement and thresholding' but lacked noise immunity. Marr and Hildreth [4] proposed Laplacian of Gaussian (LoG) operator which also had some limitations such as: considerable effect of noise on the performance of edge detection, detection of false edges etc. Canny [5] proposed another gradient based edge detector making use of Gaussian filter for smoothening of image and noise suppression. This technique removed genuine high frequency edge features while smoothing out the noise, thus degrading localization and neglecting the detection of low- contrast edges. Gudmundsson et al. [6] used genetic algorithm based on optimization for performing edge detection which proved to be successful in detecting thin, continuous and well localized edges. But, still it had certain short comings such as: big computational time complexity and low computational efficiency. The manuscript presented by Zahurul and Zahidul [7] employed improved Sobel operator for performing edge detection in orthopedic images. But the detected edges were not superior and the need of developing an enhanced refining operator had been cited. Fengjing et al. [8] proposed a color edge detection technique based on the concept of triangle similarity using an improved Sobel operator. However, the technique proved to be inefficient for noisy images. Pie et al. [9] in their work introduced 'Generalized Radial Hilbert Transform' for performing edge detection which showed higher noise immunity because of its longer impulse response. Reduction in impulse response improved the edge map quality but resulted in decrement in noise immunity. Gopalyegani et al. [10] presented a technique employing Hilbert matrix and its inverse for performing edge detection which gave inadequate results for images corrupted with impulse noise. Herein, order of both: Hilbert matrix as well as inverse Hilbert matrix had to be computed manually which was a tedious task. Martin et al. [11] utilized the concept of artificially plotted edges, compared and analyzed the role that different features play in edge detection by unsupervised learning algorithms, and combined the features to detect edges. The proposed method gave satisfactory performance; however, it needed the artificial edges as the learning object to determine the simulation results. Hu et al. [12] employed bilateral filtering and adaptive tone mapping for edge detection and its enhancement. This technique was a success but in absence of an optimum threshold, degradation of the edge map quality took place. Panetta et al. [13] introduced the "Parameterized Logarithmic Image Processing" (PLIP) model with a view to process images with more accuracy. However, this method had problems in areas that involved crossing over of regions thus damaging the original image's symmetry. Nair et al. [14] presented an edge detection algorithm employing Hyperbolic Tangent Filter (HBT) along with Logarithmic Transform which was a modified version of Luminance and Contrast-Invariant Edge Similarity Measure algorithm. The algorithm proposed in their work exhibited improvement in the SSIM values when compared with the previous work. However, the results were not appreciable in noisy and contrast variant images and thus the effectiveness of the algorithm was limited to noiseless and luminance variant images. Thus, it is seen that although current literature is abundant in a variety of edge detection algorithms, they do not always lead to acceptable results in noisy environment. Moreover, most of the techniques focused primarily on grayscale images while very few concentrated on edge detection in color images. As a remedy to this problem an approach for performing color edge detection using Hilbert transform is proposed in this paper. In order to show the effectiveness of the proposed approach in noisy environment, image corrupted with impulse noise, Gaussian noise and their mixture has been used to carry out simulations. Concept of RGB color triangle helps in maintaining the originality of image by avoiding the transformation of one color space into another. Bilateral filtering is utilized prior to the application of Hilbert transform which improves the performance of edge detection in noisy environment as it preserves edges while smoothing the image. Reconstruction estimation function has been used in this paper to perform the quality assessment of the obtained edge maps. The remaining part of this paper is structured as follows: a detailed explanation of the proposed edge detection approach has been given in Sect. 2. Performance evaluation based on the simulated results has been performed in Sect. 3. Conclusions, summarized on the basis of obtained results are drawn in Sect. 4.

2 Proposed Edge Detection Approach

The procedure employed by the proposed edge detection approach consists of four main functional blocks namely: Noise superimposition, Pre-processing, Bilateral filtering and finally edge enhancement as shown in the block diagram given in Fig. 1. In order to show the efficiency of the proposed edge detection technique in noisy environment, the input color image *p(x)* is superimposed by either Gaussian

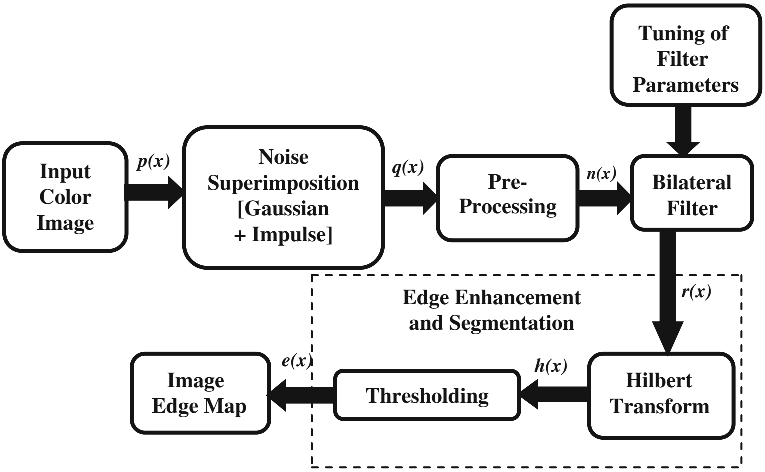


Fig. 1 Block diagram of proposed edge enhancement and segmentation approach

noise or impulse noise or mixture of impulse and Gaussian noise. This whole process comes under noise superimposition functional block. Under pre-processing functional block, the transformation of the noise inflicted image *n(x)* is done using RGB color triangle [8] resulting in a transformed image *q(x)*. The transformed image *q(x)* undergoes bilateral filtering for necessary smoothening and noise suppression yielding the filtered image *r(x)*. This step is followed by application of Hilbert transform for edge enhancement giving *h(x)* as the resultant image. For further segmentation, thresholding is performed on enhanced edges yielding the final edge map *e(x)*.

2.1 Pre-processing (Image Transformation Using RGB Color Triangle)

The pre-processing deals with the transformation of the input color image *p(x),* by mapping each of its pixels to a RGB color triangle [8]. The sides of the RGB color triangle are assigned the R, G and B values of the corresponding pixels. Perimeter of the RGB scalene triangle is calculated using (1) and stored in a separate matrix. Thus, each pixel in the original color image is replaced by the perimeter of its corresponding RGB color triangle. The process is repeated for each and every pixel of the input image finally yielding the transformed image *q(x).*

$$q(x) = aR + bG + cB \quad (1)$$

where: a, b and c are constants and have been assigned a value equal to (1/3). This pre-processing step avoids transformation of one color space into another and thereby assisted in maintaining the originality of image. The transformed image *q(x)* is then smoothened by using bilateral filtering.

2.2 Bilateral Filtering

Impulse noise, also known as Salt and Pepper noise is a special type of data dropout noise represented by a sprinkle of bright and dark spots just like salt and pepper granules. This noise usually corrupts the image during transmission, sensor faults or due to ac power interference. Gaussian noise is statistical noise that has its probability density function equal to that of the normal distribution, which is also known as the Gaussian distribution. In other words, the values that the noise can take on are Gaussian-distributed. Edges represent points in an image where there is a sharp change of intensity from one value to another. Hence, edges are present in the high frequency structure of the image and are more susceptible to noise making edge detection unstable and inaccurate leading to detection of false edges or at times loss of true edges. Gaussian low pass filter and Median filters are commonly used for image noise suppression but during this process, these filters suppress the high frequency structure leading to the removal of finer details along with noise. In the present work, bilateral filtering [15, 16] is used as an edge preserving technique to catalyze the performance of proposed edge detector in noisy environments. In bilateral filtering, each pixel x in the input image *q(x)* is replaced by the normalized weighted average of the pixels present in the spatial neighborhood *N(x)* and can be expressed mathematically as,

$$r(x) = \frac{\sum_{a \in N(x)} W(x) q(a)}{\sum_{a \in N(x)} W(x)} \tag{2}$$

where: a signifies the nearby pixel location in *N(x)* and *r(x)* represents the image reconstructed after bilateral filtering. The weight *W(x)* in (2) is calculated by the multiplication of two other weight components as given in (3).

$$W(x) = W_S(x).W_R(x) \tag{3}$$

where: $W_S(x)$ and $W_R(x)$ can be stated as:

$$W_S(x) = e^{-\frac{1}{2}\left(\frac{||a-x||}{\sigma_d}\right)^2} \tag{4}$$

$$W_R(x) = e^{-\frac{1}{2}\left(\frac{||q(a)-q(x)||}{\sigma_r}\right)^2} \tag{5}$$

In (4), $\|a–x\|$ represents the Euclidean distance between a and x and in (5), $\|q(a) - q(x)\|$ measures the intensity difference between two points a and x. The parameters σ_d and σ_r represent the geometric spread and the photometric spread respectively. The component, $W_s(x)$ in (3) performs filtering in spatial domain acting as *domain filter* whereas $W_R(x)$ performs filtering in intensity domain acting as *range filter*. The output image after this stage is $r(x)$ and the Hilbert transform is applied on this image. The role of domain filter is mainly noise suppression and smoothening whereas range filter assists in preserving crisp edges. Thus, bilateral filter can be interpreted as a combination of these two filters acting simultaneously on the input image. The performance of this filter can be tuned with the help of spread parameters σ_d and σ_r respectively depending upon the noise content present in an image.

2.3 Hilbert Transform

Livadas and Constantinides [17] were the first to present the idea of performing edge detection using Hilbert transform. This transform can be considered as a filter which shifts phases of all the frequency components of its input signal by $-\pi/2$ radians. Hilbert transform of a single dimension signal, $r(x)$ can be stated as:

$$h(x) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{r(\tau)}{\tau - x} d\tau \tag{6}$$

The algorithm for computing Hilbert transform (in two-dimension) for digital images has been given under Table 1. In the algorithm, N represents total no. of pixels in the input image.

Table 1 Algorithm to find 2- dimensional Hilbert transform for digital images

BEGIN	
Step 1:	*Matrix P(x)* ← *row wise FFT of input image r(x).*
Step 2:	*If, i=1 or (N/2)+1*
	then Q(i)=1.
	else
	if i=2, 3, …, (N/2)
	then Q(i)=2.
	else
	if i=(N/2)+2, …, N
	then Q(i)=0.
Step 3:	*Φ (x)* ← *store element wise product of P(x) and Q(x) in a matrix.*
Step 4:	*H(x)= inverse FFT of Φ(x).*
Step 4:	*Repeat step 1-3 column wise on H(x) to find the Hilbert transformed image h(x).*
END	

For segmentation of features, a suitable threshold selection is needed; this is achieved by calculating T as the average of the absolute values of pixel intensities in $h(x)$ as given by Eq. (8).

$$T = \frac{1}{N}\sum_{x=1}^{N} |h(x)| \tag{8}$$

Finally, thresholding is performed on $h(x)$ based on the condition given in (9) to generate the edge map $e(x)$.

$$[h(x) - h(x+1)] > T \tag{9}$$

3 Results and Discussion

3.1 Performance Evaluation

Many techniques have been proposed so far for evaluating the performance of edge detectors but they have specific criteria that restrict them to be used for all types of images. For instance, Pratt's Figure Merit (*PFOM*) [18] is used for the quality evaluation of edge maps for real images where as reconstruction estimation function [19] is used as a parameter for real (edge maps) images. In the present work, the evaluation of edge maps is performed under reconstruction estimation function. In this, an image is reconstructed from the obtained edge map using the pixel intensities in the grayscale version ($b(x)$) of the original image. The first step of this process is marked with the dilation of the obtained edge map $e(x)$. Pixels in the reconstructed image $z(x)$ are set equivalent to the corresponding pixel value of $b(x)$ when $e(x)$ is equal to one. This is followed by a linear interpolation until all the pixels attain an interpolated value. The reconstructed image $z(x)$ must minimize the function given by Eq. (10).

$$\int \left(\frac{\partial z}{\partial x}\right)^2 dx \tag{10}$$

The error between the reconstructed and the original image helps in assessing the quality of the generated edge map. This is attained by calculating Mean Squared Error (*MSE*) as given by Eq. (11). Lower the value of *MSE*, better is the performance of edge detector for real images.

$$MSE = \frac{1}{N}\sum_{x=1}^{N} (z(x) - b(x))^2 \tag{11}$$

3.2 Simulation Results

A 'Lena' image (real image) are taken as test image for simulations in this work. Simulations are carried out on noise free as well as noisy versions of this image. Three datasets of noisy image have been taken in this work to carry out simulations namely- Data set 1: gaussian noise (with variance $(\sigma) = 0.005$–0.02) inflicted image, data set 2: impulse noise (5–20 %) inflicted image and data set 3: image corrupted by mixture of impulse and gaussian noise. During pre-processing, each of the pixels in *p(x)* is incremented by one and then transformed using (1). The pre-processed image is then given as input to a bilateral filter with σ_d ranging between 1 and 3 and σ_r varying from 10 to 120. The values of the spread parameters are tuned according to the noise content in the image. With the application of Hilbert transform and thresholding, final edge maps are generated. For the purpose of comparison, edge detection is performed on the above set of test images using Sobel operator [3], improved Sobel operator [8], generalized radial Hilbert transform [9] along with the proposed approach. Figure 2 shows simulation results for 'Lena' image in noise free as well as all three data sets of noisy images. The obtained edge maps clearly show the efficiency of the proposed approach over other edge detection techniques. For the improved Sobel operator, the detected edges are quite thick but the performance starts getting deteriorated with increasing variance of Gaussian noise. The edge map gets degraded further in presence of mixture of impulse and guassian noises. Sobel operator misses some true edges while performing edge detection in noise free images. Moreover, the edge maps obtained by the application of Sobel operator and generalized radial Hilbert transform are not favorable in noisy environment. If the simulation results of the proposed approach are analyzed, the it is found that favorable edge maps are obtained not only in noise free image but also in noisy environment. Tabulation of *MSE* values (reconstruction estimation) for edge maps all three data sets of noisy 'lena' image has been done in Tables 2, 3 and 4 respectively. Obtained values of *MSE* are least for the edge maps generated with the proposed approach in all three cases. In addition, there is a relative increment in *MSE* values with the increase in the noise intensity. But, even for maximum noise contamination, the obtained *MSE* with the proposed technique is least in comparison to other techniques.

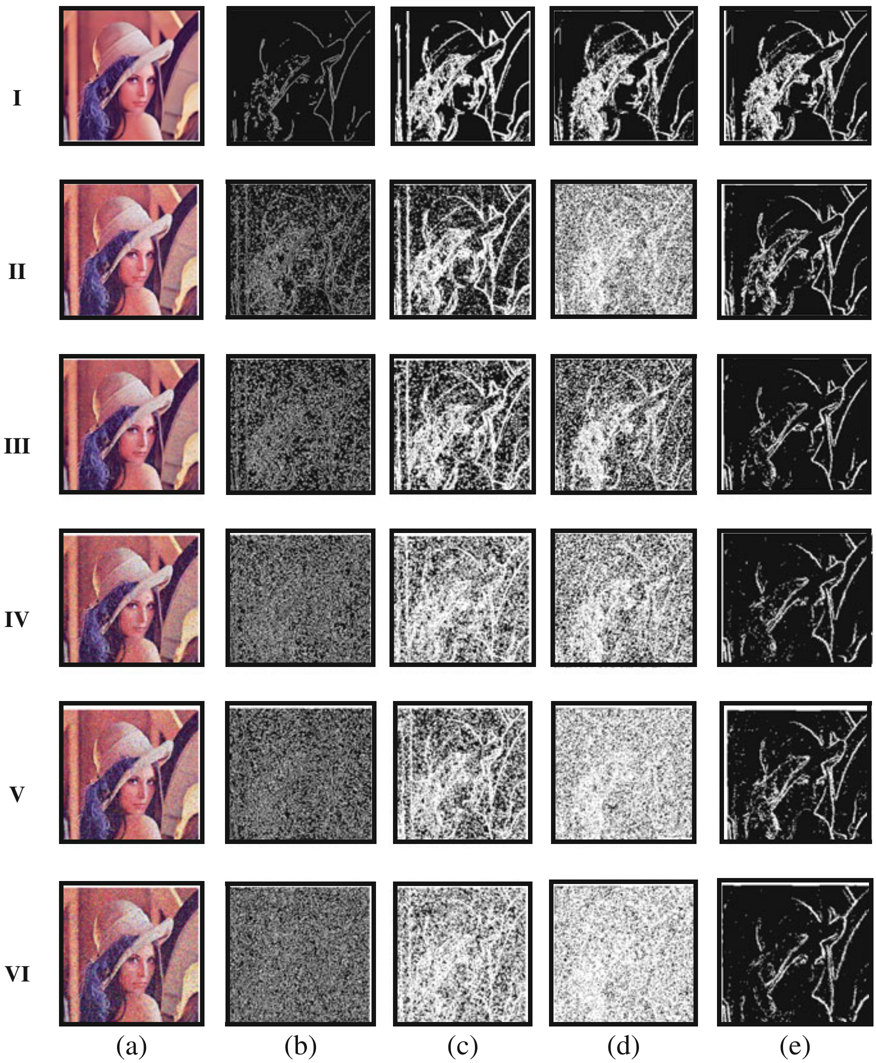


Fig. 2 **a** Original image: *I* Noise free, *II* contaminated with gaussian noise of variance = 0.02, *III* Contaminated with 5 % of impulse noise, *IV* contaminated with 10 % impulse noise, *V* contaminated with mixture of impulse (5 %) and gaussian (variance = 0.005) *VI* contaminated with mixture of impulse (10 %) and gaussian (variance = 0.01). Edge map obtained by: **b** Sobel operator [3]. **c** Improved sobel operator [8] **d** Generalized radial hilbert transform [9]. **e** Proposed technique

Table 2 *MSE* values for different edge detection techniques calculated for the 'lena' image corrupted with gaussian noise

Variance of gaussian noise	Sobel operator [3]	Improved sobel operator [8]	Generalized radial hilbert transform [9]	Proposed technique
0 (Noise free)	0.0152	0.0052	0.0050	0.0047
0.005	0.0163	0.0071	0.0131	0.0053
0.01	0.0182	0.0098	0.0211	0.0057
0.015	0.0203	0.0115	0.0238	0.0062
0.020	0.0254	0.0153	0.0317	0.0076

Table 3 *MSE* values for different Edge Detection Techniques calculated for the 'lena' image corrupted with impulse noise

Noise contamination (%)	Sobel operator [3]	Improved sobel operator [8]	Generalized radial hilbert transform [9]	Proposed technique
5	0.0179	0.0083	0.0142	0.0060
10	0.0196	0.0099	0.0199	0.0081
15	0.0223	0.0128	0.0236	0.0096
20	0.0299	0.0164	0.0310	0.0122

Table 4 *MSE* values for different edge detection techniques calculated for the 'lena' image corrupted with mixture of impulse noise and gaussian noise

Noise Contamination (impulse + gussian)	Sobel operator [3]	Improved sobel operator [8]	Generalized radial hilbert transform [9]	Proposed technique
5 % + $\sigma = 0.005$	0.0177	0.0091	0.0167	0.0063
10 % + $\sigma = 0.01$	0.0210	0.0121	0.0231	0.0087
15 % + $\sigma = 0.015$	0.0269	0.0164	0.0292	0.0102
20 % + $\sigma = 0.02$	0.0355	0.0211	0.0386	0.0136

4 Conclusion

The approach used in this paper for performing edge detection in color images has been proved to be effective in both noise free and noisy environment. The inclusion of bilateral filtering along with Hilbert transform has played a crucial role in increasing the efficiency of the proposed edge detection approach in noisy environment. The concept of RGB color triangle has played an important role in maintaining the originality of the input color image. The *MSE* values calculated under reconstruction estimation function also supports the simulation results carried out on 'lena' image.

References

1. Ziou D, Tabbone S (1998) Edge detection techniques: an overview. Int J Patt Recogn Image Anal 8:537–559
2. Novak CL, Shafer SA (1987) Color edge detection. In: Proceedings of the DARPA image understanding workshop, vol 1. pp 35–37
3. Sharifi M, Fathy M, Mahmoudi MT (2002) A classified and comparative study of edge detection algorithms. In: Proceedings of the international conference on information technology: coding and computing, pp 117–220
4. Marr D, Hildreth EC (1980) Theory of edge detection. Proc R Soc Lond B Biol Sci 207:187–217
5. Canny JF (1986) A computational approach to edge detection. IEEE Trans Patt Anal Mach Intell PAMI-8(6):679–698
6. Gudmundsson M et al (1998) Edge detection in medical images using a genetic algorithm. IEEE Trans Med Imaging 17(3):469–474
7. Zahurul S, Zahidul S (2010) An Adept edge detection algorithm for human knee osteoarthritis images. In: Proceedings of international conference on signal acquisition and processing pp 375–379
8. Fengjing Z, et al. (2012) Color image edge detection arithmetic based on color space. In: Proceedings of the international conference on computer science and electronics engineering, IEEE, pp 217–220
9. Pei S, et al. (2003) The generalized radial hilbert transform and its applications to 2-d edge detection (any direction or specified direction). In: Proceedings of the international conference on acoustics, speech and signal processing, pp 357–360
10. Golpayegani N, et al. (2010) A novel algorithm for edge enhancement based on hilbert matrix. In: Proceedings of 2nd international conference on computer engineering and technology, vol 1. pp 579–581
11. Martin DR et al (2004) Learning to detect natural image boundaries using local brightness, color, and texture cues. IEEE Trans Patt Anal Mach Intell 26(5):530–549
12. Hu KJ, et al. (2009) Bilateral filtering and adaptive tone mapping for qualified edge and image enhancement. In: Proceedings of SPIE-IS&T electronic imaging, vol 7241. pp 1–8
13. Panetta K, et al. (2007) Parameterization of logarithmic image processing models. IEEE Trans Syst Man Cybern Part A: Syst Hum 1–12
14. Nair MS, et al. (2009) A novel approach for edge detection using HBT filter and logarithmic transform. In: Proceedings of international conference on digital image processing, pp 243–246
15. Tomasi C, Manduchi R (1998) Bilateral filtering for gray and color images. In: Proceedings of the IEEE international conference on computer vision, Bombay, India, pp 839–846
16. Elad M (2002) On the origin of bilateral filter and ways to improve it. IEEE Trans Image Process 11(10):1141–1151
17. Livadas GL, Constantinides AG (1988) Image edge detection and segmentation based on the hilbert transform. In: Proceedings of the international conference on acoustics, speech and signal processing, pp 1152–1155
18. Pratt WK (1991) Digital image processing, 2nd edn. Wiley, New York
19. Agaian SS et al (2010) Boolean derivatives with application to edge detection for imaging systems. IEEE Trans Syst Man Cybern Part B: Cybern 40(2):371–382

Classification of ECG Images Using Probabilistic Neural Network Based on Statistical Feature Analysis

R. Nandhini and P. Subhasini

Abstract Research on the analysis of ECG is mostly used for automating the diagnosis of different cardiac diseases. The ECG waveforms may differ for same patient to such extent that they are unlike each other and at same time alike for different types of beats. Many algorithms have been developed for the classification and detection of the ECG beat. In order to improve the accuracy of the ECG image feature extraction and classification system, the present research work proposes the use of different feature extraction methods. ECG image feature extraction and classification system, uses five feature extraction methods, namely wavelet decomposition, edge detection, gray level histogram, fast fourior transform and mean–variance, and probabilistic neural network classification. The objective of this present research work is to achieve the high accuracy and simplest classifiers related to extract the input features. An ECG image is classified by PNN using various feature extraction. The experimental results shows that wavelet decomposition gives a maximum accuracy compared to other feature extraction methods (Tayel MB, El-Bouridy ME (2008) ECG images classification using artificial neural network based on several feature extraction methods. International conference on computer engineering and system pp 113–115).

Keywords Wavelet decomposition (WT) · Edge detection (ED) · Fast fourier transform (FFT) · Gray level histogram (GLH) · Mean–variance (MV) · Probabilistic neural network (PNN).

R. Nandhini (✉) · P. Subhasini
Department of Computer Science, Avinashilingam Institute for Home Science and Higher Education for Women University, Coimbatore 600043, India
e-mail: nandhu2k9@gmail.com

P. Subhasini
e-mail: p.subashini@gmail.com

Mohan S. and S. S. Kumar (eds.), *Proceedings of the Fourth International Conference on Signal and Image Processing 2012 (ICSIP 2012)*, Lecture Notes in Electrical Engineering 222, DOI: 10.1007/978-81-322-1000-9_50,

1 Introduction

Electrocardiogram is originated from the action of human heart. It is mainly used for diagnosing cardiac diseases. The basis of ECG is the electrical activation of heart muscle cell causing depolarization of its membrane. The depolarization is propagated along the cell fiber and transmitted to adjoining cells. ECG can measure the rate and rhythm of the heartbeat and it provides indirect evidence of blood flow to the heart muscle. 12 lead electrode systems are used for ECG recording and give an overall view of heart's electrical activity. ECG analysis system depends on the reliable and accurate detection of the PQRST waves along with the measurement of QT segments. Different approaches are used to develop more accurate algorithms for extracting the features. The proposed algorithm is divided into three steps. The first step is reading the ECG images, which converts the ECG images into numerical data, each ECG image has dimension of (512By648) pixels.

Then convert input images from RGB format to gray scale. After converting to gray images, feature extraction is one of the most important steps in the classification system, since the feature selection has to represent well the pattern which has to be recognized [1]. In many case, the amount of data selected in the feature extraction is huge so the reduction of this data is necessary.

2 Methodology

The primary objective of the present research work is to extract feature values from ECG images. Five feature extraction methods were selected namely, wavelet decomposition, edge detection, fast fourier transform, gray level histogram and mean–variance. As shown in Fig. 1, the proposed system divided into three steps.

This research methodology consists of three stages.

Stage 1. The input images are transforming from RGB to gray
Stage 2. Feature values are extracted from ECG images
Stage 3. Final step is classification process, and calculating the accuracy

2.1 Reading the ECG Image

In stage 1, read the used ECG images which convert the 50 diseases ECG images into numerical data, each image has dimension of (512By648). Then convert the input images from RGB to gray images. Figure 2, shows the original and the resultant gray ECG images

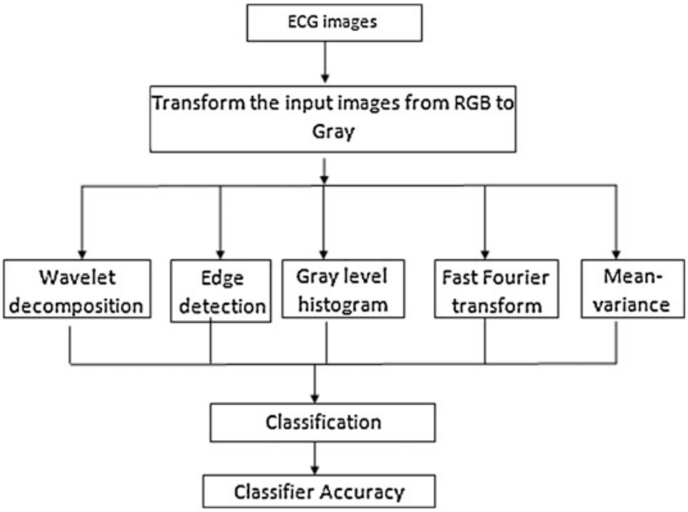


Fig. 1 The proposed system

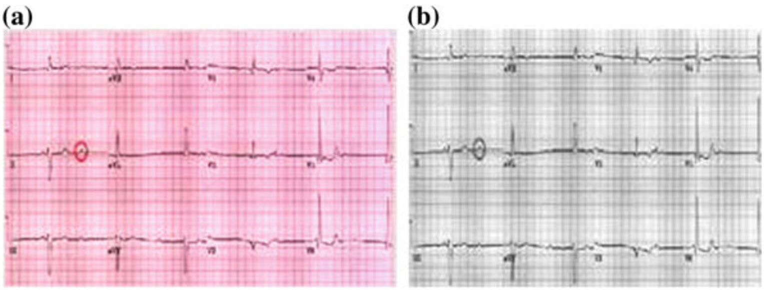


Fig. 2 RGB to Gray image. **a** Original image. **b** RGB to gray image

2.2 Feature Extraction

Feature extraction is a special form of dimensionality reduction in pattern recognition and image processing, machine learning and data mining. It involves simplifying the amount of resources required to describe a large set of data accurately. Transforming the input data into set features is called feature extraction. In this research work, five feature extraction methods are used namely, wavelet decomposition, edge detection, fast fourier transform and mean–variance.

2.2.1 Wavelet Decomposition

It is powerful tool for decomposing the original signal or images into a combination of a set of function with different location and scales. The image is decomposed into two groups of frequency. The frequencies of waveforms are high frequency (wavelet functions) and low frequency of images (scaling function). There are two types of wavelet transform mostly used namely, discrete wavelet transform (DWT), continuous wavelet transform (CWT). The discrete wavelet transform is used to analyze the signal at different resolution through the decomposition of the signal into several successive frequency bands [2]. The DWT utilizes two set of functions ¢ (t) and f (t), each associated with the low pass and the high pass filters respectively [3].

The proposed method is based on the two dimensional wavelet packets transform. The two dimensional wavelet transform for image application leads to a decomposition of approximation at level j in four components are shown in Fig. 3 the approximation at j + 1, and the details in three ordinations(horizontal, vertical, diagonal). The input signal is decomposed into two sub band signals, which is called a discrete wavelet, transform of order two [1].

The two dimensional wavelet transform can be applied to the coarser version at half the resolution, recursively, in order to further decor related neighboring pixels of the input image. The sub bands in the next higher transform levels l will be denoted by LL (l), LH (l), HL (l), and HH (l), where LL (1) = LL, LH (1) = LH, HL (1) = HL, and HH (1) = HH, respectively. The image is decomposed into four groups of components, namely, low frequency components and high frequency components in the y-direction, x-direction and xy-direction.

2.2.2 Edge Detection

Edge detection is a fundamental tool used in most image processing applications to obtain information from images as a precursor step to feature extraction. In computer vision, edge detection is a fundamental task that needs to point out the true edges to get the best results depends on the performance of this task. The edge

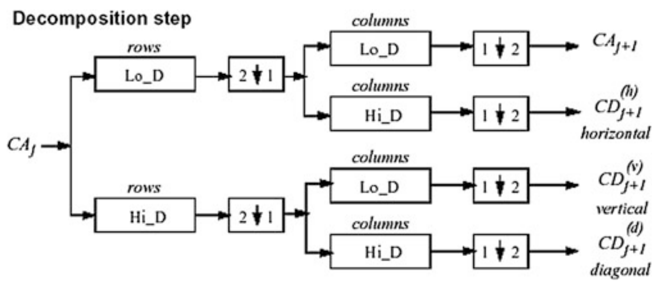


Fig. 3 Two dimensional wavelet transformation

detection aims to localize the boundaries of objects in an image, it is a basis for many image analysis and machine vision applications. Conventional approaches to edge detection are computationally expensive because each set of operations is conducted for each pixel. In conventional approaches, the computation time quickly increases with the size of the image. The edge detection is based on the behavioral study of these edges with respect to the following differentiation operators:

- Canny edge detector
- Sobel edge detector
- Prewitt edge detector
- Laplacian of Gaussian
- Robert edge detector

The proposed system is based on the sobel edge detection method.

Sobel Edge Detection

The Sobel operator is used in image processing, particularly within edge detection algorithms. Technically, it is a discrete differentiation operator, computing an approximation of the opposite of the gradient of the image intensity function. At each point in the image, the result of the Sobel operator is either the corresponding opposite of the gradient vector or the norm of this vector [4]. The Sobel operator is based on convolving the image with a small, separable, and integer valued filter in horizontal and vertical direction and is therefore relatively inexpensive in terms of computations. On the other hand, the opposite of the gradient approximation that it produces is relatively crude, in particular for high frequency variations in the image [5].

In simple terms, operator calculates the opposite of the gradient of the image intensity at each point, giving the direction of the largest possible change from light to dark and the rate of change in that direction. Mathematically, the operator uses two 3×3 kernels which are convolved with the original image to calculate approximations of the derivatives—one for horizontal changes, and one for vertical [6]. The sobel edge detection masks look for edges in both the horizontal and vertical directions and then combine this information into a single metric.

In this proposed system, sobel edge detection has more efficiency than comparing with other edge detection. Edge detections usually based on the calculation of intensity gradient across the images. There are four feature extractions values that are calculated and stored in a vector at column shape with length of equal elements. The extracted features using edge detection method for all diseases images are used as input to the PNN classifier [1].

2.2.3 Fast Fourier Transform

The Fast Fourier Transform (FFT) is the powerful tool for analyzing and measuring signals from plug-in data acquisition (DAQ) devices. The Fourier transform is used in a wide range of applications, such as image analysis, image filtering, image reconstruction and image compression. Fast Fourier transform computes the discrete Fourier transform and it produces same results as evaluating the DFT definition directly.

Fast Fourier transform is used to convert the images from frequency domain to spatial domain through vice versa. The frequency domain used to change the image position corresponding to the changes in the spatial frequency. The mean, median, standard deviation, variance of the 50 ECG images are calculated and stored respectively in matrix (4 by 50) elements. The extracted features using FFT method for all 50 ECG images are used as input to the PNN classifier.

2.2.4 Gray Level Histogram

The gray level histogram is used in the proposed method. A grayscale (or gray level) image is simply one in which the only colors are shades of gray. The reason for differentiating such images from any other sort of color image is that less information needs to be provided for each pixel. The gray-scale histogram of an image represents the distribution of the pixels in the image over the gray-level scale. It can be visualised as, if each pixel is placed in a bin corresponding to the colour intensity of that pixel. All of the pixels in each bin are then added up and displayed on a graph. The extracted features using Gray level histogram method for all 50 ECG images are used as input to the PNN classifier.

2.2.5 Mean–Variance

The most common method in statistical distribution is mean; with a discrete random variable is the mathematical average of all the terms. To calculate it, add up the values of all the terms and then divide by the number of terms. This expression is also called the arithmetic mean. There are other expressions for the mean of a finite set of terms but these forms are rarely used in statistics [7]. The mean of a statistical distribution with a continuous random variable, also called the expected value, is obtained by integrating the product of the variable with its probability as defined by the distribution. The expected value is denoted by the lowercase Greek letter mu (μ). Mean of a data set is simply the arithmetic average of the values in the set, obtained by summing the values and dividing by number of values.

The *variance* of a data set is the arithmetic average of the squared differences between the values and the mean. The extracted features using Mean–variance method for all 50 ECG images are used as input to the PNN classifier.

3 ECG Classification

Classification is most important in image processing, which is also known as pattern recognition, discrimination, supervised learning or prediction. It is a task that involves construction of a procedure that maps data into one of several pre-defined classes [8]. In the present work, Probabilistic neural network is used. Probabilistic neural networks are feed forward networks built with three layers. They are derived from Bayes decisions networks. Probabilistic neural networks estimate the probability density function for each class based on the training samples. PNN uses parzen or a similar probability function, this is calculating for each test vector. Vector must be normalized prior to input into the network [9]. There is an input unit for each dimension in the vector. The input layer is fully connected to the hidden layer. The hidden layer has a node for each classification. Each hidden node calculates the dot product of the input vectors with a test vector that subtracts 1 from it and divides the result by the standard deviation squared. The output layer has a node for each pattern classification [10]. The sum for each hidden node is sent to the output layer and the highest values.

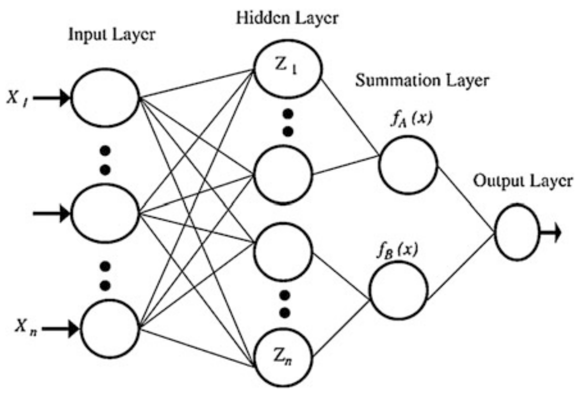
In PNN algorithm, calculating the class-node activations is a simple process. For each class node, the example vector activations are summed, which are the sum of the products of the example vector and the input vector. The hidden node activation, shown in the following equation, is simply the product of the two vectors (E is the example vector, and F is the input feature vector).

The architecture of probabilistic neural networks is shown in Fig. 4.

4 Result and Discussion

ECG Image classification is an area of research which has proved to be challenging for the past several decades and it has gained more attention in ECG image feature extraction due to the new challenges posed by ECG images. In this research work,

Fig. 4 Architecture of PNN



feature extraction ethics are constructed, which takes different input features from the images. These features are used to train and test the classifiers. The various statistical features are extracted namely, mean, standard deviation, variance and median. This research work proposes five feature extractions which use these features for the classification task. In present work 50 diseases ECG images are selected from the physiobank and EMEDU-database for analyzing and reconigition. The ECG record contains 10 different types of abnormal heartbeats. They are, Atrial fibrillation, atrial flutter, complete right block, AV block, acute myocardial injury, atrioventricular dissociation, tachycardia, Wolff syndrome, supra tachycardia, supraventricular tachycardia. Image features namely, mean, standard deviation, median, variance. The ECG images dataset consists of 50 images, out of which (60 %) were taken as training data and (40 %) images were taken as testing data.

The performance metric used to evaluate the proposed classification models is accuracy. To analyze the efficiency of the proposed algorithms, the results of the proposed classification systems are compared with their feature extraction algorithms. The feature extraction methods like Wavelet decomposition, Edge detection, Fast Fourier transform, Gray level histogram and mean–variance are all classified using PNN classifier to find the best feature extraction technique.

Training and testing data sets of ECG images were used to check the classification performance and its accuracy that is given by,

$$\text{Accuracy} = \frac{\text{No. of correctly detected heart diseases}}{\text{Total no. of tested heart diseases}} \times 100 \tag{4.1}$$

From the proposed system the accuracy calculated for five feature extraction techniques using probabilistic neural network classification.

The classification results showed that the accuracy of the PNN classifier using features obtained from the five preprocesses as shown in Table 1 and Fig. 5.

Table 1 Comparison of PNN classifier accuracy

Classification accuracy	
Methods	Accuracy (%)
Wavelet decomposition	93.33
Edge detection	53.33
Fast fourier transform	80
Gray level histogram	73.33
Mean–Variance	60

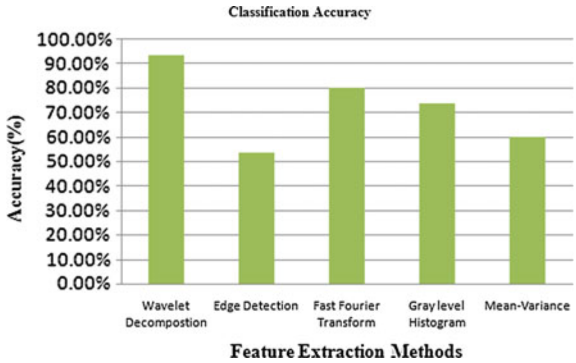


Fig. 5 PNN classification accuracy

5 Conclusion

A supervised classifier based on Probabilistic neural network is proposed for the identification and classification of gray scale digital ECG images by several feature extraction methods. The input to the PNN classifier include extracted feature using wavelet decomposition, edge detection, fast fourier transform, gray level histogram and mean–variance methods. The proposed system can identify and classify the ECG diseases images based on the five feature extraction techniques. The result of the classification process leads to that, the proposed PNN classifier using wavelet decomposition features as input gives about 93.3 % classification accuracy. This feature extraction method can be used along with a classifier computerized diagnosis of many heart diseases without depending on the experience of cardiologist, and before sending the patient to the cardiology.

References

1. Tayel MB, El-Bouridy ME (2008) ECG images classification using artificial neural network based on several feature extraction methods. International conference on computer engineering and system pp 113–115
2. Tayel MB, El-Bouridy ME (2006) ECG images classification using feature extraction based on wavelet transformation and neural network. 6th International conference on computer engineering and system, pp 105–107
3. Martis RJ, Chakraborty C, Ray AK (2009) An integrated ECG feature extraction scheme using PCA and wavelet transform. Indian conference on digital object identifier, pp 1–4
4. Subashini P, Jansi S (2011) A study on detection of focal cortical dysplasia using MRI brain images. J Comp Appl JCA, vol 4, issue 2, ISSN Number: 0974-1925, pp 23–28
5. 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
6. Wu J, Yin Z, Xiong Y (2007) The fast multilevel fuzzy edge detection of blurry images. IEEE Signal Process Lett 14(5):344–347
7. Srinivasan GN, Shobha G (2008) Statistical texture analysis. Proceedings of world academy of science, engineering and technology, vol 36

8. Ibrahiem MM, Emary El, Ramakrishnan S (2008) On the application of various probabilistic neural networks in solving different pattern classification problems. World applied sciences journal, pp 772–780
9. Rutkowski L (2004) Adaptive probabilistic neural networks for pattern classification in time-varying environment. IEEE Trans Neural Netw (15) pp 811–827
10. Chen Y-H, Yu S-N (2006) Comparison of different wavelet subband features in the classification of ECG beats using probabilistic neural network. 28th annual international conference on engineering medicine and biology society pp 1398–1401
11. Kim T, Yang HS (2006) A multidimensional histogram equalization by fitting an isotropic gaussian mixture to uniform distribution. International conference on image processing, Atlanta
12. Banerjee S, Mitra M (2010) ECG feature extraction and classification of anteroseptal myocardial infarction and normal subjects using discrete wavelet transform. International conference on systems in medicine and biology, pp 55–60
13. Ĩnesta JM, Calera-Rubio J (2002) Robust gray-level histogram gaussian characterization. Caelli T et al (eds) SSPR&SPR 2002, LNCS 2396, pp 833–841

A Novel HVS Based Image Contrast Measurement Index

Mridul Trivedi, Anupam Jaiswal and Vikrant Bhateja

Abstract Image quality assessment based on HVS has become an important means to evaluate the degree of enhancement provided by contrast manipulation algorithms. In this paper a perceptual quality evaluation model using no-reference approach is proposed for estimation of contrast in digital images. The proposed index is formulated using the operators of LIP model for measurement of contrast using two rectangular windows around the centre pixel called foreground and background. It can be visualized from the experimental results that the proposed index is capable to assess the performance of different contrast manipulation algorithms in comparison to other methods. In addition, the results of subjective analysis show that the assessment is well correlated with known characteristics of HVS.

Keywords Contrast measurement index · No-reference · HVS · Measure of enhancement · Local contrast

M. Trivedi (✉) · A. Jaiswal · V. Bhateja
Department of Electronics and Communication Engineering, Shri Ramswaroop Memorial Group of Professional Colleges, Faizabad Road, Lucknow 227105 UP, India
e-mail: mridultrivedi1991@gmail.com

A. Jaiswal
e-mail: anupam4.srmc@gmail.com

V. Bhateja
e-mail: bhateja.vikrant@gmail.com

Mohan S. and S. S. Kumar (eds.), *Proceedings of the Fourth International Conference on Signal and Image Processing 2012 (ICSIP 2012)*, Lecture Notes in Electrical Engineering 222,
DOI: 10.1007/978-81-322-1000-9_51,

1 Introduction

Image quality assessment in digital domain is critical in all applications of image processing because when an image is transformed, the viewer is the ultimate judge of how well a transformation method works. Therefore, HVS becomes a major platform for the measurement of enhancement [1]. Image enhancement algorithms seek to enhance the apparent visual quality of an image or emphasize certain features based on the knowledge of source of degradation. An important feature that influences image enhancement is contrast, as it is a perceptual measure that defines the difference between the perceived brightness [2]. Objective evaluation methods use the mathematical expressions for assessment of image quality while the subjective methods are based on physiological and psychological properties of human. They evaluate the quality of the image as perceived by an individual observer [3]. Objective evaluation methods for digital images are preferred as they can dynamically monitor and adjust image quality, optimize parameter settings and can be used to benchmark various digital images processing systems [4]. However, it is also a necessary condition that the results of objective evaluation must be coherent with perceived visual quality based on HVS characteristics [1]. Objective methods of quality evaluation are of three types: Full-Reference, Reduced-Reference, and No-Reference. In full-reference methods, a processed image is compared to its original one by calculating the difference between corresponding pixels in two images found at the same pixel locations. Mean Squared Error (MSE) [4] and Structural Similarity (SSIM) [5] are the two most commonly used full-reference signal fidelity measures. MSE exhibits weak performance and has been widely condemned for serious shortcomings as it does not confirm to the expected results when it is used to predict human perception of assessing image quality. On the other hand, SSIM cannot handle geometrical distortions in digital images which are non-structural in nature. To overcome the drawbacks of above methods, reduced-reference evaluation methods [6] were introduced which involve supplying some amount of information such as statistical parameters about the reference images along with the distorted image which may be useful in quality computation. Reduced-reference methods are widely criticized for not correlating well with the perceived image quality. On the other hand, no-reference methods [7] do not require any reference images for their assessment rather a blind evaluation is performed based on some characteristics of the given image. On the basis of assessment of image quality, these measures return an absolute value which is generally content dependent and calculation is based on the specific types of distortions. These types of assessment methods require reference images for quality evaluation. Hence we need No-reference evaluation methods for the assessment of contrast enhancement.

2 Literature Review

2.1 No-Reference Methods for Image Contrast Evaluation

Several methods of image contrast evaluations for no-reference images have been proposed. Tripathi et al. in their work [8] proposed Histogram Flatness Measure (HFM) and Histogram Spread (HS) measures for evaluation of enhancement on the basis of statistical parameters of image histogram like geometric mean, quartile distance and range. However, it was concluded by the authors that both the measures worked satisfactorily well only for evaluation of histogram based enhancement methods and that too with reduce sensitivity. Measure of Enhancement (EME) and the Measure of Enhancement by Entropy (EMEE) are common no-reference methods for evaluation of image contrast proposed by Agaian et al. [9]. These methods evaluated the image quality by computing the ratio of local maximum and minimum pixels in small $k_1 \times k_2$ sized blocks and the results are averaged for the entire image. To measure the similarity of images, Wharton et al. [10] found that the EME could be used as an image similarity measure, called SEME. To assess image similarity, a modified EME measure was used to assess the quality of an image. The SEME measure was able to assess image similarity across a range of distortions in a better way. This is used to measure the entropy of the error signal (the difference between the images) by calculating the absolute values of the minimum and maximum luminances. But these methods rely on linear algorithms and can assess satisfactorily only when there is a large background with small single test object. LogAME and LogAMEE were the two modifications of the previous methods developed by incorporating a non-linear framework to the Michelson Contrast Law [11]. As these methods work on small blocks in an image so the results are affected by noise and steep edges in images [12]. To overcome this problem, a new enhancement measure using the concept of the second derivative (SDME) [13] is introduced since it measures the change ratio of the variation speed of pixel. In this method, the image is divided into $k1 \times k2$ blocks and then the maximum and minimum value of the pixels including the intensity of centre pixel in each block is calculated separately. This metric is not very useful for real world luminance values because of the logarithmic response characteristic of human eye [13]. Hence, this paper presents a novel HVS [1] based image contrast measurement index using non-reference approach. The proposed index incorporates LIP model and evaluates the contrast using two rectangular windows around the centre pixel called foreground and background. The proposed index evaluates the contrast manipulated by enhancement algorithms on test images taken from standard MATLAB library as well as LIVE Database [14]. Subjective analysis shows the robustness of the proposed index and its coherence with human evaluation.

2.2 Logarithmic Image Processing Model

The Logarithmic Image Processing (LIP) model of Jourlin and Pinoli [15] is a mathematical framework that provides a specific set of non-linear algebraic and functional operations for the processing and analysis of pixel intensities. The LIP model has been proved to be physically justified by some important laws and characteristics of human brightness perception. This is designed to both maintain the pixels values inside the range as well as for more accurate processing of images from a HVS [1] point of view. The LIP operations are defined using gray tone functions expressed as $k(i, j)$ in (1).

$$k(i,j) = m - f(i,j) \tag{1}$$

where: $f(i, j)$ is the original image and m denotes the maximum value of the pixel in that image. $k(i, j)$ in (1) is the gray tone function used to generate negatives of the original images. Addition and subtraction using LIP operators can now be expressed in term of gray tone functions as follows:

$$k_1 \mathbin{\widetilde{\oplus}} k_2 = k_1 + k_2 - \frac{k_1 k_2}{m} \tag{2}$$

$$k_1 \mathbin{\widetilde{\Theta}} k_2 = m \frac{k_1 - k_2}{m - k_2} \tag{3}$$

where: $\widetilde{\oplus}$ and $\widetilde{\Theta}$are operators for LIP addition and subtraction respectively. k_1 and k_2 represents the corresponding gray tone functions. Instead of processing the pixels with basic arithmetic operations, LIP arithmetic operator's yields more robust functioning well within the acceptable range of image. The remaining part of this paper is organized as follows: Sect. 3 describes the Proposed Image Contrast Measurement Index in context to discussion of basic definition of contrast. The simulation results and their discussion are given in Sect. 4. Section 5 concludes the paper.

3 Proposed Image Contrast Measurement Index

Contrast has a great influence on the quality of an image in human visual perception as well as in image analysis. The definition of local contrast proposed by Morrow et al. [2] can be stated as:

$$C = \frac{m_f - m_b}{m_f + m_b} \tag{4}$$

where: m_f is the maximum luminance equivalent to the mean gray level value of a particular object in the image called the foreground. In the same context, m_b equals minimum luminance which is the mean gray level value of region surrounding that object, called the background. If the difference in the intensities between foreground and background is more than 2 % then change in contrast cannot be properly distinguished by human eye. With this concept, at times the measurement of contrast using (4) yields poor sensitivity. Moreover, the variation of contrast provided by certain enhancement algorithms Histogram Equalization (HE) [16], Adaptive Histogram Equalization (CLAHE) [17], Unsharp Masking (UM) [18], Adjusting the Black to White Mapping (ABWM) [19], Morphological operations such as Bottom Hat filtering by square (BHS) [20] structuring elements were not quantized sharply when evaluated using the above expression. The results are also not coherent as per HVS characteristics. Further, this contrast measurement approach is not versatile as the value of contrast varies with the selection of foreground and background regions. Hence, there is a need of a quality index which can effectively measure contrast in digital images overcoming the limitations discussed shortcoming as well as capable to discriminate between increasing and decreasing contrast in accordance to HVS [1]. This paper proposes a contrast measurement index for assessment of various contrast manipulation algorithms, using the operators of LIP model. The degree of improvement in contrast provided by an enhancement algorithm can be adjudged by the fact that it should enhance the difference between the average gray level values lying in the foreground and background regions respectively. On this basis, the algorithm for computation of proposed quality index is given in Table 1.

Table 1 Algorithm for computation of proposed quality index for contrast measurement

BEGIN

Step 1: *The centre o(x, y) of the input image, I of size $r \times c$ is determined as $(r + c)/2$ (where r, $c \in$ odd numbers).*

Step 2: *Taking o(x, y) as centre, two concentric square windows (of sizes 3×3 and 5×5) are selected. The smaller window of size 3×3 can be referred to as the 'foreground' whereas the larger window of size 5×5 is called background as shown in Fig. 1a. Mean gray level values of these foreground and background regions can be computed and denoted as M_f and M_b respectively.*

Step 3: *The contrast within this region can be calculated as (5) in a sub-region of the input image during the first iteration.* $C_i = \ln\left|\left(\frac{M_f \tilde{\Theta} M_b}{M_f \tilde{\oplus} M_b}\right)_i\right|$ (5)

Step 4: *Keeping the centre fixed at o(x, y), the foreground and background window sizes are incremented by a factor of 2 and respective values of mean gray levels is computed for both foreground and background regions. Hence, during the second iteration, mean gray level values are calculated for foreground and background regions of window sizes 5×5 and 7×7 respectively as shown in Fig. 1b. This can be stated as:* $C_{i+1} = \ln\left|\left(\frac{M_f \tilde{\Theta} M_b}{M_f \tilde{\oplus} M_b}\right)_{i+1}\right|$ (6)

Step 5: *The procedure is iteratively repeated for the entire image, by incrementing the foreground and background window sizes, each time by a factor of 2, till the time the background window size reaches r.*

Step 6: *The contrast value computed during each iteration are averaged for all the iterations and multiplied by a factor α*

END

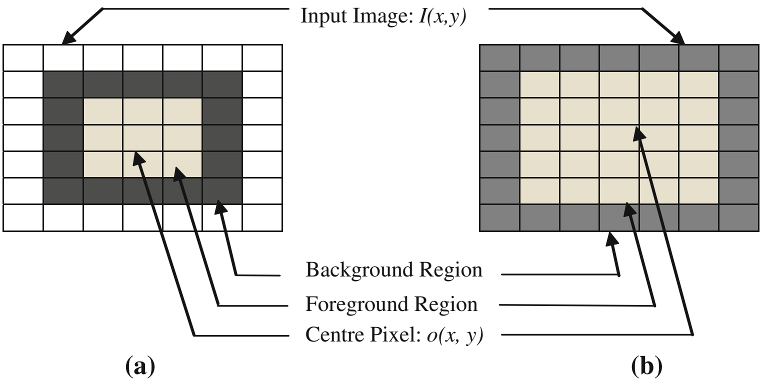


Fig. 1 Selection of foreground and background regions for contrast evaluation during the first two iterations. **a** Shows the 3 × 3 foreground and 5 × 5 background window selected about the centre pixel: *o(x, y)*. **b** Shows the increment in window sizes to previous ones, with 5 × 5 foreground and 7 × 7 background windows about the same centre pixel: *o(x, y)*

α in the above algorithm is equivalent to the maximum of the Hadamard Transform computed for the original image I. Hence, the proposed quality index known as Contrast Measurement Index (*CMI*) can be mathematically formulated as:

$$CMI = \frac{\alpha}{N}\sum_{i=1}^{N} \ln\left|\left(\frac{M_f \,\widetilde{\Theta}\, M_b}{M_f \,\widetilde{\oplus}\, M_b}\right)_i\right| \tag{7}$$

where: N is the total no. of iterations applied for contrast evaluation. Equation (7) yields an absolute value for the input image (I) which is a measure of its contrast. Higher values of *CMI,* characterizes better performance of the contrast enhancement algorithms. The range of *CMI* can be bounded between zero and infinity where zero indicates a totally black image and infinity for totally white image. The absorption and transmission of light follows a logarithmic relationship, both when processed by the human eye and when travelling through a medium [10] because of this it is natural to combine the statistical parameters of contrast measurement using logarithmic operators. This justifies the usage of LIP model to ensure HVS based contrast evaluation.

4 Experimental Results

4.1 Test Images

To test the versatility of the proposed quality index number of experiments were included on two categories of test images: standard MATLAB images (1, 2 and 3) and LIVE Database images (4, 5 and 6) as shown in Fig. 2a. The test images are initially processed by converting it from RGB to gray-scale and then normalized, prior to the application of contrast manipulation algorithms. Certain algorithms are applied for increasing the contrast such as Histogram Equalization (HE) [16], Adaptive Histogram Equalization (CLAHE) [17], Unsharp Masking (UM) [18] as given in Fig. 2d, whereas the contrast is decreased by adjusting the black to white mapping (ABWM) [19], applying Morphological operations such as Bottom Hat filtering with a square (BHS) [20] as shown in Fig. 2e, f.

4.2 Simulation Results

The proposed index *CMI* is calculated for all the test images in Fig. 2 and tabulated under Table 2. From the data in Table 2, it can be interpreted that the value of *CMI* increases with respect to its value for the original image, upon application of increasing contrast algorithms [16–18]. Similarly, a decrease in value of *CMI* (in comparison to original image) is seen with decreasing contrast algorithms [19, 20]. This illustrates that the proposed index is capable to discriminate between

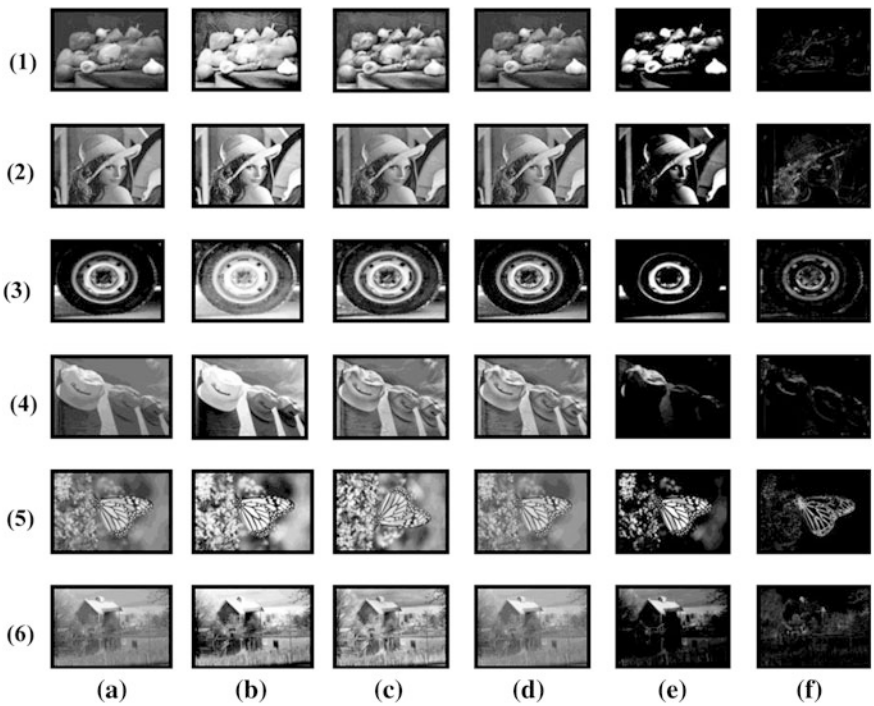


Fig. 2 Different contrast manipulation algorithms applied on test images (1–6) **a** original test images, (1–3): from MATLAB library and (4–6): from LIVE database. Images transformed by **b** HE [16], **c** CLAHE [17], **d** UM [18]. **e** ABWM [19], **f** BHS [20]

Table 2 Calculation of *CMI* for different contrast manipulation algorithms

Image No.	Original Image	HE [16]	CLAHE [17]	UM [18]	ABWM [19]	BHS [20]
(1)	4.3211	6.1171	6.4295	6.6219	3.4612	2.4123
(2)	14.3211	19.7244	19.7318	17.3581	12.3124	3.1265
(3)	5.5989	9.8221	7.1893	5.8116	3.2345	3.6748
(4)	6.2376	8.1233	8.3326	8.6831	4.7376	2.1723
(5)	11.5987	15.3462	14.3949	19.5886	6.8894	3.9534
(6)	13.2383	16.3650	15.5038	17.9331	5.2123	3.1243

images of good and poor contrast without the knowledge of original reference images. The sensitivity of *CMI* is better as there is sharp change in its value with application of different contrast manipulating algorithms. HE [16] enhances the contrast better than other algorithms as it equalizes the histogram of the image uniformly, resulting in an image consisting of gray-levels with density, increasing the dynamic range of the image. This can be clearly verified by proposed index as it evaluates to a maximum *CMI* value for images transformed using HE [16] among various enhancement algorithms. CLAHE [17] uses a clip level to limit the local histogram such that the amount of contrast enhancement for each pixel can be limited. Hence, there is always a control on over-enhancement was performed with HE. Therefore, values of *CMI* for images transformed using CLAHE are less in magnitude in comparison to those with HE. For the purpose of comparisons, the contrast of the test images is also calculated using previously proposed well known measures of enhancements, namely EMEE [7] and LogAME [10] and the values are given in Table 3. It can be observed that, these evaluation methods proved to be non-satisfactory in discriminating good and poor contrast images. There is an increase in the values of both EMEE and LogAME for images transformed with decreasing contrast algorithms [19, 20]. Not only this, the sensitivity of LogAME also comes out to be very low in comparison to the proposed index. UM [18]

Table 3 Calculation of EMEE and LogAME for different contrast manipulation algorithms

	Image No.	Original Image	HE [16]	CLAHE [17]	UM [18]	ABWM [19]	BHS [20]
EMEE	(1)	3.9094	26.8221	39.1207	52.6201	82.3145	111.6736
	(2)	13.9961	59.3921	29.5745	59.9370	47.3213	142.6595
	(3)	39.2501	39.6500	71.3668	94.3620	99.3315	102.4314
	(4)	54.2234	22.4013	73.7529	73.1239	97.2316	121.2134
	(5)	42.7891	91.0390	78.7148	58.9332	101.2376	132.3737
	(6)	33.8617	29.6600	77.2662	58.0487	114.2350	147.3370
LogAME	(1)	0.0015	0.0013	0.0017	0.0011	0.0017	0.1213
	(2)	0.0718	0.0683	0.0665	0.0648	0.0442	0.0442
	(3)	0.0013	0.0013	0.0012	0.1571	0.0811	0.1563
	(4)	0.2341	0.0712	0.0711	0.0181	0.0823	0.0441
	(5)	0.0520	0.0491	0.0492	0.0197	0.0401	0.4077
	(6)	0.0511	0.0462	0.0463	0.0645	0.0422	0.0489

Table 4 Correlation coefficients for different contrast evaluation methods

Contrast Evaluation methods	PC	SRC
CMI	0.8801	0.8494
EMEE	−0.6676	−0.5977
logAME	−0.1346	−0.0897

approach generates a high contrast image, by adding some portion of the original image to the blurred image. The obtained values of EMEE for images transformed using UM are not as stable as *CMI* values.

4.3 Subjective Evaluation

In this paper the efficiency of the *CMI* is assessed by the subjective evaluation [16] of images in Fig. 2. To perform subjective analysis, images (1–6) are transformed by various contrast manipulating algorithms. They are evaluated by a group of 60 observers. The quality scores given by the observers were recorded on a quality scale of 1–100 (1-poor quality and 100-best quality). The Mean Opinion Score (MOS) for each image is calculated and their values are normalized between 0 and 1. Scores obtained by using *CMI,* EMEE and LogAME are also normalized to obtain the correlation of their values with respect to MOS.

The correlation can be calculated by using Pearson's (PC) and Spearman's Rank Correlation (SRC) and tabularized in Table 4. It can be visualized from Table 4 that the proposed index CMI yields the best correlation result with HVS. EMEE [7] shows negative correlation as it gives unstable result on decreasing contrast algorithms. The result of LogAME [10] does not support HVS [1], as its correlation factor is quite low. Figure 3 shows the scatter plot diagram between MOS and objective methods. EMEE [7] varies along a straight line with MOS but with a negative slope while logAME [10] shows no correlation with MOS as circle are scattered all over the plot. However, in *CMI* the circles in scatter plot are along a straight line with positive slope which shows that *CMI* has good correlation with HVS.

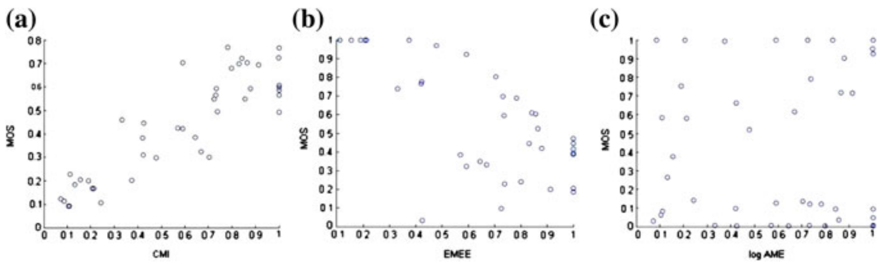


Fig. 3 Scatter plots of different contrast indices with MOS **a** CMI **b** EMEE **c** LogAME

5 Conclusion

Common measures of enhancement like EME, EMEE, LogAME etc. are well known for their non-reference evaluation approach, computational simplicity and applicability for optimization purposes. However this somehow yields non satisfactory performance as the assessment result is not coherent with HVS characteristics. Unlike the earlier proposed methods of contrast evaluation, *CMI* evaluates the local contrast by calculating the difference in the mean gray-level values of foreground and the background regions using the operators of LIP model. It can be inferred from the simulation results that *CMI* is more robust and provides an improved evaluation of contrast based on HVS characteristics. Subjective evaluation provided a strong base for assessing the efficiency of the proposed work according to the human perception.

References

1. Wharton E, Panetta K, Agaian S (2008) Human visual system based similarity metrics. Proc IEEE Conf Syst Man Cybern 17(3):685–690
2. Morrow WM et al (1992) Region-based contrast enhancement of mammograms. IEEE Trans Med Imaging 11(2):121–134
3. Seshadrinathan K, Soundararajan S, Bovik AC, Cormack LK (2010) Study of subjective and objective quality assessment of video. In: IEEE Trans Image Process 19(6):1427–1441
4. Sheikh HR, Saber MF, Bovik AC (2006) A statistical evaluation of recent full reference image quality assessment algorithm. IEEE Trans Image Process 15(11):3441–3452
5. 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
6. Wang Z, Simoncelli EP (2005) Reduced-reference image quality. In: Proceedings of international symposium on electronic imaging, San Jose, CA, USA
7. Agaian SS, Panetta K, Grigoryan AM (2001) Transform-based image enhancement algorithms with performance measure. IEEE Trans Image Process 10(3):367–382
8. Tripathi AK, Mukhopadhyay S, Dhara AK (2011) Performance metrics for image contrast. In: IEEE conference on image information processing, Shimla, India
9. Panetta K, Wharton EJ, Agaian SS (2008) Human visual system based image enhancement and logarithmic contrast measure. IEEE Trans Image Process 38(1):174–188
10. Wharton E, Agaian SS, Panetta K (2006) A logarithmic measure of image enhancement. In: Proceedings of SPIE, defense security system, Orlando, pp 620–629
11. Peli E (1990) Contrast in complex images. J Opt Soc Am 7:2032–2040
12. Agaian SS, Silver B, Panetta K (2007) Transform coefficient histogram-based image enhancement algorithms using contrast entropy. IEEE Trans Image Process 16:741–758
13. Panetta K, Zhou Y, Agaian SS, Jia H (2011) Nonlinear unsharp masking for mammogram enhancement. IEEE Trans Inf Technol Biomed 15(6):234–255
14. Sheikh HR, Wang Z, Cormack L, Bovik AC (2005) LIVE image quality assessment database release 2. http://live.ece.utexas.edu/research/quality
15. Jourlin M, Pinoli JC (2001) Logarithmic image processing; the mathematical and physical framework for the representation and processing of transmitted images. Adv Imaging Electron Phys 115:129–196

16. Chen SD, Ramli AR (2003) Contrast enhancement using recursive mean-separate histogram equalization for scalable brightness preservation. IEEE Trans Consumer Electron 49(4):1301–1309
17. Zuiderveld K (1994) Contrast limited adaptive histogram equalization. In: Graphic Gems IV. Academic Press Professional, San Diego, pp 474–485
18. Ramponi G, Polesel A (1998) Rational unsharp masking technique. J Electron Imaging 7:333–338
19. Gonzalez RC, Woods RE (2002) Digital image processing. Addison-Wesley, Reading
20. Boomgard V et al (1992) Methods for fast morphological image transforms using bitmapped images. Comput Vis Graph Image Process Graph Models Image Process 54(3):252–254

Efficient Speaker Independent Isolated Speech Recognition for Tamil Language Using Wavelet Denoising and Hidden Markov Model

C. Vimala and V. Radha

Abstract Current research on Automatic Speech Recognition (ASR) focuses on developing systems that would be much more robust against variability in environment, utterance, speaker and language. In this paper all these major factors are considered to develop a system which works powerfully for recognizing a set of Tamil spoken words from a group of people at different noisy conditions. Developing an ASR system in the presence of noise critically affects the speech quality, intelligibility, and recognition rate of the system. Thus, to make a system robust against different noisy conditions, the most popular speech enhancement techniques such as spectral subtraction, adaptive filters and wavelet denoising are implemented at four SNR dB levels namely −10, −5, 5 and 10 with three types of noise such as white, pink and babble noise. This research work is carried out for developing a speaker independent isolated speech recognition system for Tamil language using Hidden Markov Model (HMM) under the above noise conditions. Better improvements are obtained when the proposed system is combined with speech enhancement preprocessor. Based on the experiments 88, 84 and 96 % of recognition accuracy are obtained from enhanced speech using Nonlinear Spectral Subtraction, RLS adaptive Filter and Wavelet approach respectively.

Keywords Nonlinear spectral subtraction · RLS adaptive algorithm · Wavelet denoising · MFCC · HMM · Tamil speech recognition

C. Vimala (✉) · V. Radha
Department of Computer Science, Avinashilingam Institute for Home Science and Higher Education for Women, Coimbatore, Tamil Nadu, India
e-mail: vimalac.au@gmail.com

V. Radha
e-mail: radhasrimail@gmail.com

Mohan S. and S. S. Kumar (eds.), *Proceedings of the Fourth International Conference on Signal and Image Processing 2012 (ICSIP 2012)*, Lecture Notes in Electrical Engineering 222,
DOI: 10.1007/978-81-322-1000-9_52,

1 Introduction

Speech recognition allows the system to identify the words that a person speaks into a microphone or telephone and convert them into written text. The biggest advantage of using ASR is the ability to achieve hands-free computing. It also offers huge benefit for people with disability who find difficulties in using a keyboard and mouse. Hence it is becoming an attractive alternate choice for users to manage applications through voice rather than a mouse or keyboard. The significant applications of ASR system include voice dialing, call routing, automatic transcriptions, information searching, data entry, speech-to-text processing and aircraft etc. Based on the requirement, the ASR system can be classified into different broad categories. Accordingly it can be classified as isolated or continuous speech and speaker-dependent or speaker-independent system. Some of the speech recognition applications require speaker-dependent system whereas some systems need speaker independent system where the inter-speaker variability should be eliminated. Based on these constraints, it is desirable in many applications requiring small well defined vocabularies which efficiently work for a group of people. Hence, today's researches are more focused on developing speaker independent isolated speech recognition systems. But, the great disadvantage of speaker-independent system is that the error rate is normally higher than speaker-dependent systems. In recent times, with the combination of more sophisticated techniques and improved independent speech recognition engines it offers excellent performance with improved productivity.

Another important aspect in speech recognition system is its performance level in noisy environment. Most of the systems achieve reliable performance in noise free environments but works very poor in noisy conditions. Developing a highly effective speech recognition system which achieves greater accuracy in noisy conditions is a challenging task [1]. Also, the greater part of these speech recognizers have been primarily developed for English language. Hence, today numerous researchers are mainly focusing on developing speech recognizers for their native languages.

The paper is organized as follows. The overview of a proposed system is given in Sect. 2 and the analysis of various speech enhancement techniques are explained in Sect. 3. The Sect. 4 deals with feature extraction using MFCC and Sect. 5 briefly explains about the HMM based speech recognition. The experimental results are shown in Sect. 6. The subjective and objective performance evaluation is given in Sect. 7. Finally, the optimum method for speaker independent isolated speech recognition system for Tamil language is concluded in Sect. 8.

2 Overview of the Proposed System

The proposed system involves various preprocessing and feature extraction techniques as front end for the ASR system. The Fig. 1 shows the overview of speaker independent isolated speech recognition system for Tamil language.

Initially, the system receives the analog signal which is converted into a digital signal using Digital Signal Processor (DSP). Next, the digitized speech is given to the speech preprocessing system which will perform dc offset removal and pre-emphasis. Subsequently, the preprocessed signal is given to the speech enhancement system for speech noise cancellation. After that, the useful feature vectors are extracted from the enhanced speech signals using MFCC. Finally, the HMM model is used to recognize the spoken word based on these feature vectors. These techniques are explained in the subsequent sections.

3 Speech Enhancement Techniques

In real time environment, speech signals are normally corrupted by numerous types of noise. The occurrence of noise in speech significantly decreases the intelligibility and the quality of the signal. Reducing noise and enhancing the perceptual quality and intelligibility of a speech without disturbing the signal quality are the vital task. Hence, speech enhancement algorithm plays a crucial role in speech recognition to improve the accuracy in noisy environment. Several techniques have been proposed for this purpose namely spectral subtraction, adaptive noise cancellation, kalman filtering, fuzzy algorithms, extended and iterative wiener filtering, HMM-based algorithms and signal subspace methods

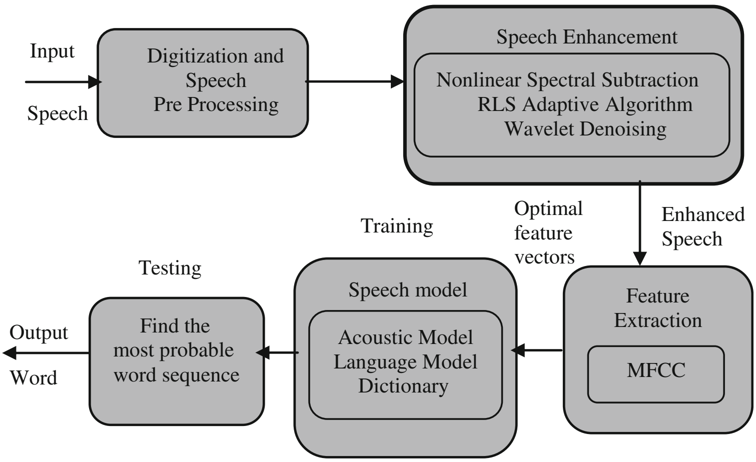


Fig. 1 Overview of Speaker independent isolated speech recognition system

etc. Achieving the high quality and intelligibility of the processed speech signal is the major target of these techniques. The main objective of speech enhancement is to improve the following:

- speech signal quality and intelligibility
- robustness of speech which is affected by noise
- speech signal-to noise ratio
- accuracy of speech recognition systems operating in noisy environments

In this paper, three types of speech enhancement techniques are implemented namely spectral subtraction, adaptive filtering and wavelet. For this work, three types of noises are considered namely white, pink and babble noise at four SNR dB levels namely −10, −5, 5 and 10. The following section gives a brief outline of these techniques.

3.1 Spectral Subtraction

The spectral subtraction is the conventional method proposed for additive background noise. It is mainly used to suppress the noise from the degraded signal and it became more popular due to its simplicity [2]. It is represented in Eq. (1)

$$y(n) = s(n) + d(n) \tag{1}$$

where y(n) is the noisy speech which is composed of the clean speech signal s(n) and the additive noise signal d(n). It is implemented by estimating the spectral magnitude during no speech and subtracting this spectral estimate of the noise from the subsequent speech spectral magnitude [3]. This technique is effective for additive noise reduction, however they introduce some artificial noise which alters the original signal. Hence these algorithms has gone through many modifications with time and introduced new methods. They include Nonlinear Spectral Subtraction (NSS), MultiBand Spectral Subtraction, Minimum Mean Square Error (MMSE), and Log Spectral MMSE [4]. In this paper, all these techniques are implemented and the performances are measured both subjectively and objectively. Based on the analysis, it was found that, the NSS is the premium method for better noise cancellation [5].

3.1.1 Nonlinear Spectral Subtraction

Among the spectral subtraction techniques, the non-linear approach gives better result since it make use of frequency dependent subtraction factor for different types of noise [6]. In this method, in order to produce improved results, larger values are subtracted at frequencies with low SNR levels and smaller values are subtracted at frequencies with high SNR levels. It is estimated using Eq. (2).

$$\begin{aligned} |X_e(\omega)| &= |Y(\omega)| - \alpha(\omega)N(\omega) \\ &\mathit{if} \\ |Y(\omega)| &> \alpha(\omega)N(\omega) + \beta|D_e(\omega)| \\ &\text{else} \\ \beta &= |Y(\omega)| \end{aligned} \tag{2}$$

where β is the spectral floor, $\alpha(\omega)$ is a frequency dependent subtraction factor and N(ω) is a non-linear function of the noise spectrum

$$N(\omega) = Max(|D_e(\omega)|) \tag{3}$$

where $N(\omega)$ is the maximum of the noise magnitude spectra.

$$\begin{aligned} \alpha(\omega) &= \frac{1}{r + P(\omega)} \\ \mathrm{y(t)} &= \mathrm{x(t)} + \mathrm{n(t)} \\ \mathrm{Y_{j,k}} &= \mathrm{X_{j,k}} + \mathrm{N_{j,k}} \end{aligned} \tag{4}$$

where r is a scaling factor and $P(\omega)$ is the square root of the posteriori SNR estimate given as

$$P(\omega) = |Y(\omega)|/|D_e(\omega)| \tag{5}$$

The performance of this algorithm is further considered for comparing with the other two enhancement techniques.

3.2 Adaptive Filtering

Adaptive filters also called self learning filters which do not have constant filter coefficients and do not require a priori knowledge of signal or noise characteristics [7]. It has the potential to achieve enhanced performance in an environment where information of the relevant statistics is not presented [8]. The well known and popular kind of adaptive filters are Least Mean Square (LMS), Normalized Least Mean Square (NLMS) and Recursive Least Squares (RLS) algorithms. The performances of all these algorithms are analyzed to find out the efficient adaptive algorithm for speech enhancement. Among these, LMS and NLMS algorithms are very simple and effective method to implement but they are slower. Whereas the RLS algorithm makes the converging speed and also offers better noise reduction and enhanced speech quality and intelligibility when compared to the other algorithms. The RLS algorithm and its advantages are discussed below.

3.2.1 Recursive Least Squares Algorithm

RLS algorithms offer excellent performance in time varying environments like speech [9]. It is a recursive implementation of the Wiener filter which is used to find the filter coefficients that relate to producing the recursively least squares of the error signal i.e. the difference between the desired and actual signal. In contrast to LMS and NLMS algorithm, the RLS aims to reduce the mean square error. At each instant, an exact minimization of the sum of the squares of the desired signal estimation errors are performed by referring the values of previous error estimations. The RLS approach offers faster convergence and smaller error with respect to the unknown system, at the expense of requiring more computations [10]. In this paper, the performance of LMS, NLMS and RLS algorithms are experienced at different noisy conditions. As a result, it was observed that the performance of RLS algorithm is superior to other adaptive algorithms [10].

3.3 Wavelet Denoising

Wavelet denoising is a non-parametric estimation method that has been proposed in recent years for speech enhancement applications. Transform domain always plays an important role in any speech signal processing application. Fourier transform was the earlier choice of domain but creates annoying musical noise in speech noise suppression [11]. Later, some methods have been proposed to solve this problem but have not achieved satisfied performance. In recent years, wavelet domain based approach has been found to be a very useful tool for solving various problems particularly for speech denoising [11]. Compared to Fourier transform, it is possible in wavelets to obtain a good approximation of the given function f by using only a few coefficients which is the great metric [12]. If an observed signal includes unwanted noise, the result is an additive signal model given by (6)

$$\mathrm{y(t)} = \mathrm{x(t)} + \mathrm{n(t)} \tag{6}$$

where $\boldsymbol{y}$ is the noisy signal, $\boldsymbol{x}$ is the clean signal, and $\boldsymbol{n}$ is the additive noise.

Then the wavelet transform performs the following (7)

$$\mathrm{Y_{j,k}} = \mathrm{X_{j,k}} + \mathrm{N_{j,k}} \tag{7}$$

where $\mathrm{Y_{j,k}}$ represents the kth set of wavelet coefficients across the selected scale j.

In wavelet transform, the noise is typically represented across time with smaller coefficients whereas signal energy is focused on larger coefficients. This improves the possibility of separating the signal from the noise based on some threshold [12]. The main advantage of wavelet denoising is that it removes noise from the corrupted signal without modifying the speech content. Among different types of wavelet family, Daubechies wavelets are a powerful and efficient approach as they are orthogonal and localized both in real and Fourier space [12]. In this paper, level 8 Daubechies 4 wavelet has been implemented which offers comparatively good performance than some other wavelets family.

All the above three algorithms are implemented and their performances are measured both subjectively and objectively. For the experiments, the separate noise corpus from NOIZEUS were collected and added to the Tamil Speech signals. Analysis is done on noisy speech signal corrupted by white, pink and babble noise at −10, −5, 5 and 10 SNR dB levels. The following Fig. 2 shows the results of speech enhancement using Nonlinear Spectral Subtraction, RLS adaptive algorithm and Wavelet approach for a sample signal corrupted by 10 dB babble noise. It is clear from the experiments that the wavelet technique offers better performance for all types of noise and SNR levels which are considered for this work. The subjective measures of these techniques are explained in the performance evaluation section. The enhanced output signals from these techniques are taken as an optimum input signal for feature extraction.

4 Speech Feature Extraction

The most significant part of all recognition systems is the feature extraction, which converts the speech signal into some digital form of meaningful features.

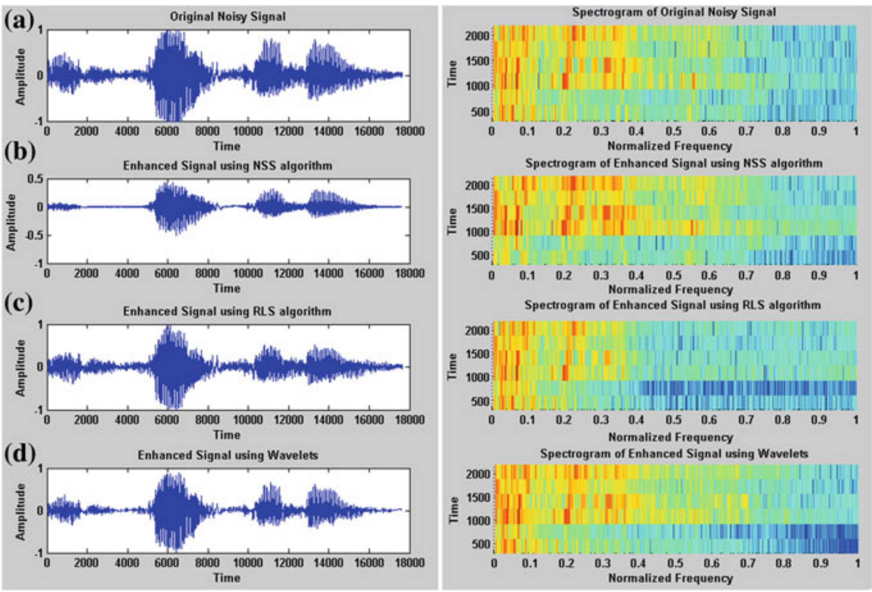


Fig. 2 Results of speech enhancement and its spectrogram representations. **a** Original babble noise signal at 10 dB SNR. **b** Enhanced signal using nonlinear spectral subtraction. **c** Enhanced signal using RLS adaptive algorithm. **d** Enhanced signal using wavelet

Providing prominent features is the major task in speech recognition system to achieve good accuracy. The good choice of features makes the speech recognition job easier in classification. The widely used feature extraction techniques are Linear Predicting Coding (LPC), Mel Frequency Cepstral Coefficient (MFCC), RASTA and Perceptual Linear Prediction (PLP). In this paper, the MFCC feature extraction technique is chosen due to its prominent characteristics. The following section gives details about MFCC feature extraction.

4.1 Mel Frequency Cepstral Coefficients

The human ear has high frequency resolution in low-frequency parts of the spectrum and low frequency resolution in the high-frequency parts of the spectrum. The coefficients of the power spectrum of a speech signal can be transformed to reflect the frequency resolution of the human ear. MFCC is based on the short-term analysis, and thus from each frame a feature vector is computed. To extract these coefficients the speech sample is taken as an input and hamming window is applied to minimize the discontinuities of a signal. Then Discrete Fourier Transformation (DFT) is used to generate the Mel filter bank. According to Mel frequency wrapping, the width of the triangular filters varies and so the log total energy in a critical band around the center frequency is included. After warping this log energy, the numbers of coefficients are obtained. This coefficient takes the best effects of Discrete Cosine Transformation (DCT) for the cepstral coefficients calculation. MFCC can be computed by using the Eq. (8)

$$\mathrm{Mel}\ (\mathrm{f}) = 2595^{*}\log 10\,(1 + \mathrm{f}/700) \tag{8}$$

Typical feature vector of 39 different parameters for each 10 ms of speech are extracted. It can be observed from the experiments that the performance of the system can be improved if the energy and delta coefficients are used additionally. These features are given as the input for the post processing steps and they are explained in the following section.

5 Speech Recognition Using HMM

Undoubtedly, modern general-purpose speech recognition systems are based on HMM. The reason for using HMM is that it can be trained automatically and it is simple and computationally feasible to use. The greatest advantage of HMM model is it is more robust to environment noise and distortions. It is a stochastic approach with set of states and transition probabilities between those states. Here, each state describes a stationary stochastic process and the changeover from one state to another to define how the process changes its characteristics in time [13].

Each state of the HMM can model the generation of the observed symbols using a stationary stochastic emission process [13]. It allows phoneme or word rather than frame-by-frame modeling of speech. Basically, the problem of speech recognition can be stated as follows. In a given acoustic observation $X = X_1, X_2 \ldots X_n$, the objective is to find out the matching word sequence $W = W_1, W_2 \ldots W_m$ which has the maximum posterior probability P(W|X) [14]. It can be defined by the Eq. (9)

$$W = \arg\max_{w} P(W/X) = \arg\max_{w} \frac{P(W)P(X/W)}{P(X)} \tag{9}$$

where P(W) is the probability of word W uttered and P(X|W) is the probability of acoustic observation of X when the word W is uttered. P (X|W) is also known as class conditioned probability distribution. P(X) is the average probability that observation X will occur [14]. The goal is to find the word W by maximizing the probability of X. To accomplish the above task an ASR system requires three significant components which are also considered as post processing steps. They are acoustic model, dictionary and language model. Since all these models are language independent, these models are created manually for Tamil language. Once all these models are developed, then the HMM model will perform likelihood evaluation, state sequence decoding and HMM estimation to find out the most probable word sequence from the given observation vector. The experimental results and the performance evaluation of the proposed system are given in the following sections.

6 Experimental Results

The main objective of this research work is to develop an ASR system to recognize isolated speech in Tamil Language from a group of speakers. Speaker independent speech recognition system itself makes the system complex. In addition to speaker characteristics, the environment, vocabulary size, Signals to Noise Ratio (SNR) also make the system very complex to implement. In this paper, all these factors are considered but the vocabulary size is limited to 50 words. Since speech corpora are not available for Tamil language it is created manually. The corpus containing 50 utterance of isolated speech were collected from 10 females. The database consists of 10 repetitions of every word produced by each speaker. The experiments were carried out in three conditions namely clean speech, noisy speech and enhanced speech with different type of noisy conditions. The best input signal for speech recognition system is identified according to the results obtained from speech enhancement. Next, the useful feature vectors extracted using MFCC are given as the input for the HMM based speech recognition. It can be observed that the HMM offers better results and higher accuracy in clean and enhanced signals. The wavelet based enhanced signals has achieved a great performance compared to Nonlinear Spectral Subtraction and RLS adaptive filters. In this work, both objective and subjective measures are considered to ensure the performance of speech enhancement techniques for improved accuracy and they are explained in the next section.

7 Performance Evaluations

Developing an ASR system continues to be a challenging field for researchers due to a great number of factors, which cause variability in speech. The main objective of this work is to recognize a set of Tamil words from a group of people under different noisy conditions. In this paper, the performance evaluations are done for both speech enhancement and recognition separately.

7.1 Evaluation of Speech Enhancement Techniques

For speech enhancement, three types of techniques are employed namely Non-linear Spectral Subtraction, RLS adaptive algorithm and wavelet based denoising approach. These algorithms are measured based on three metrics namely Mean Square Error (MSE) Rate, Peak Signal to Noise Ratio (PSNR) and Signal to Noise Ratio (SNR). Based on these measures it was found that the wavelet based denoising method confirms its superiority by offering less MSE and higher SNR and PSNR values. The wavelet method has produced better signal quality and intelligibility in all the noisy conditions and at different SNR dB levels which are implemented in this research work. The performance of wavelet was found to be comparatively good with subjective listening tests also.

7.2 Evaluation of Speech Recognition Accuracy

The speech recognition system is generally measured in terms of Word Error Rate (WER), which is the ratio between misclassified words, and total number of tested words. Generally, ASR research has been focused on minimizing the recognition error to zero in real-time independent of vocabulary size, noise, speaker characteristics and accent. The proposed system can able to recognize 48/50 words on an average by using clean signal. The following Figs. 3 and 4 illustrate the performance of the proposed system based on accuracy rate and WER for different types of signal employed in this research work.

The average accuracy of a given database is shown in Fig. 3 and it is clear from the above figures that the wavelet approach with HMM performs extremely well for all the datasets at different conditions for all the speakers enrolled in the database. The results prove that the proposed system offered higher accuracy level and less WER.

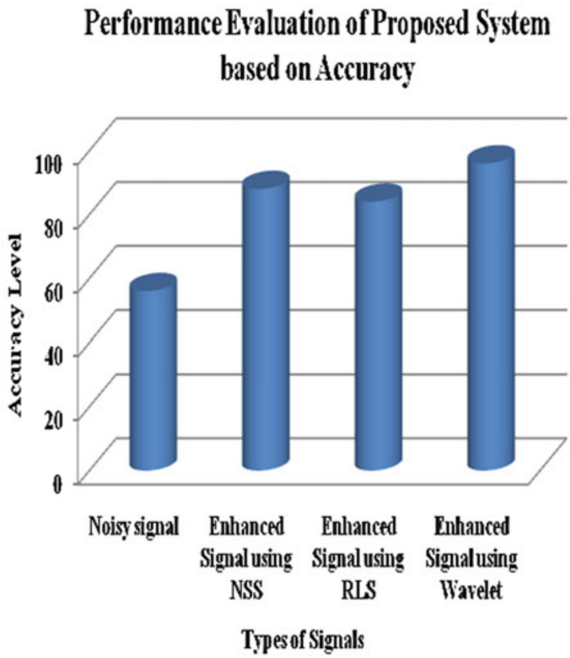


Fig. 3 Performance evaluation of proposed system based on accuracy

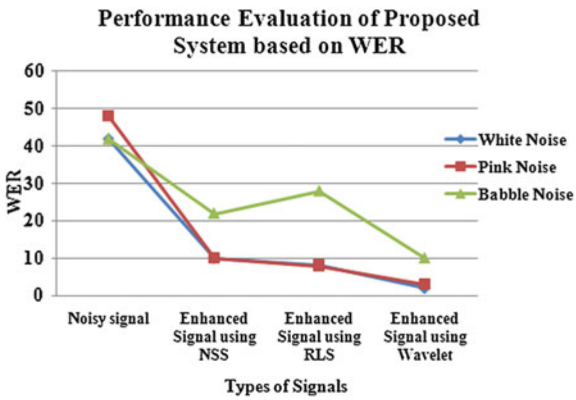


Fig. 4 Performance evaluation of proposed system based on WER

8 Conclusion

In recent years, considerable amount of research works were focused on human computer interaction through speech recognition system. Whereas language barrier is one of the significant factors which make these technologies less accessible. Based on this specific need, in this paper an efficient speaker independent isolated speech recognition system is implemented for Tamil language. The proposed system is developed using HMM since it is a most flexible and successful approach to speech recognition. The main objective of this work is to recognize the spoken words under different noisy conditions. In this paper, three types of noises are used namely white, pink and babble noise at four SNR dB levels like −10, −5, 5 and 10. The most popular speech enhancement techniques namely spectral subtraction, adaptive filters and wavelet denoising are implemented for removing the above mentioned noise types. The performances of these techniques were evaluated based on MSE, PNSR and SNR values. Based on the results, it was found that the spectral subtraction algorithm has achieved good noise cancellation but fails to produce intelligibility in enhanced speech. The RLS adaptive algorithm also performed well and offered good results in speech quality and intelligibility to some extent. Compared to the above algorithms, very good performance was achieved through wavelet denoising for all the noise factors considered for this work. The prominent input obtained from these speech enhancement methods are further analyzed with speech recognition accuracy. The comparison is done with different conditions like clean, noisy and enhanced signals. Comparatively, excellent accuracy and minimum WER are obtained through wavelet approaches. The system is found to be successful as it can identify spoken word even in noisy conditions which is the greatest advantage of HMM. Based on the experiments, it is concluded that the combination of wavelet and HMM approach can yield a highly effective recognition performance for speaker independent isolated speech recognition for Tamil language.

References

1. Baker JM, Deng L, Khudanpur S, Lee C-H, Glass J, Morgan N (2006–2007) Historical development and future directions in speech recognition and understanding. MINDS. Report of the speech understanding working group
2. Krishnamoorthy P, Mahadeva Prasanna SR (2009) Temporal and spectral processing methods for processing of degraded speech: a review. IETE Tech Rev 26(2):137–148
3. Fukane AR, Sahare SL (2011) Different approaches of spectral subtraction method for enhancing the speech signal in noisy environments. Int J Sci Eng Res 2(5). ISSN 2229-5518
4. Goel1 P, Garg A (2012) Developments in spectral subtraction for speech enhancement. Int J Eng Res Appl (IJERA). 2(1):055–063. ISSN: 2248-9622
5. Vimala C, Radha V (2012) A family of spectral subtraction algorithms for tamil speech enhancement. Int J Soft Comput Eng (IJSCE) 2(1). ISSN: 2231-2307
6. Lockwood P, Boudy J (1992) Experiments with a nonlinear spectral subtractor (NSS) hidden markov models and the projection for robust speech recognition in cars. Speech Commun 11(2–3):215–228
7. JaganNaveen V, Prabakar T, Venkata Suman J, Devi Pradeep P (2010) Noise suppression in speech signals using adaptive algorithms. Int J Signal Process Image Process Pattern Recogn 3(3):87–96
8. Hadei SA, Student member IEEE, Lotfizad M (2010) A family of adaptive filter algorithms in noise cancellation for speech enhancement. Int J Comput Electr Eng 2(2):1793–8163
9. Borisagar KR, Kulkarni GR (2010) Simulation and comparative analysis of LMS and RLS algorithms using real time speech input signal. Global J Res Eng 10(5):44 (Ver1.0)

10. Vimala C, Radha V (2012) Optimal adaptive filtering technique for tamil speech enhancement. Int J Comput Appl (0975–8887) 41(17):23–29
11. Chavan MS, Chavan MN, Gaikwad MS (2010) Studies on implementation of wavelet for denoising speech signal. Int J Comput Appl (0975–8887) 3(2):1–7
12. Johnson MT, Yuan X, Ren Y (2007) Speech signal enhancement through adaptive wavelet thresholding. Speech Commun 49(2):123–133
13. Lama P, Namburu M (2010) Speech recognition with dynamic time warping using MATLAB. CS 525, SPRING 2010—Project report
14. Thangarajan R, Natarajan AM, Selvam M (2008) Word and triphone based approaches in continuous speech recognition for tamil language. WSEAS Trans Signal Process 4(3). ISSN: 1790-5022

Prediction Accuracy of BPN by Levenberg–Marquardt Algorithm for the Prediction of COD from an Anaerobic Reactor

R. Vijayabhanu and V. Radha

Abstract Anaerobic wastewater treatment differs from traditional aerobic treatment where no aeration is used. In this paper, a model is built using Back Propagation Neural Network (BPN) that analyzed the data from an anaerobic reactor containing the cheese dairy wastewater. Data preprocessing is a crucial task to identify the efficient parameters that contribute best solution in reducing the training time with high accuracy. Data preprocessing includes data cleaning, data transformation and data reduction. The data set is screened using k-means clustering and normalized using statistical normalization techniques viz., z-score, minmax, biweight, tanh and double sigmoidal. Among the adopted normalized techniques z-score normalization produced satisfactory prediction results in BPN. The z-score normalization eases the training of BPN to predict the value of Chemical Oxygen Demand (COD) for a concentrated cheese-dairy wastewater from an anaerobic reactor. The normalized dataset is trained with BPN using Levenberg–Marquardt algorithm. The performance of the above model is evaluated based on Mean Square Error (MSE) and Regression Coefficient R. The prediction results were close to the observed data and the model was found satisfactory with MSE = 0.53375 and R = 0.99185.

R. Vijayabhanu (✉) · V. Radha
Avinashilingam Institute for Home Science and Higher Education for Women—University, Coimbatore, India
e-mail: vijayabhanuresearch@gmail.com

V. Radha
e-mail: radhasrimail@gmail.com

Mohan S. and S. S. Kumar (eds.), *Proceedings of the Fourth International Conference on Signal and Image Processing 2012 (ICSIP 2012)*, Lecture Notes in Electrical Engineering 222,
DOI: 10.1007/978-81-322-1000-9_53,

1 Introduction

Wastewater may be defined as a combination of the liquid or water-carried wastes removed from residences, institutions, and commercial and industrial establishments, together with such groundwater, surface water and storm water as may be present. The disposal of large quantity of wastewaters without adequate treatment results in significant environmental pollution. This has made researchers and scientists to have major attention in developing Wastewater Treatment systems. However, serious environmental and public health problems may result from improper operation of a wastewater treatment system (WWTS), as discharging contaminated effluent to a receiving water body can cause or spread various diseases to human beings [1]. WWTS involves several complex physical, biological and chemical processes. Often these processes exhibit non-linear behaviors which are difficult to describe by linear mathematical models [2]. In addition, the variability of the influent characteristics, in terms of composition, strength and flow rates, might influence model parameters and consequently operational control, significantly [3]. Anaerobic digestion techniques have low initial and operational costs, smaller space requirements, high organic removal efficiency and low sludge production, combined with a net energy benefit through the production of biogas. Anaerobic digestion provides an effective way of degrading organic material in wastewater [4]. The organic pollutants are converted by anaerobic microorganisms to gas containing methane and carbon dioxide, known as "biogas" [5]. Various anaerobic reactor design have been developed over the past two decades namely anaerobic filter, upflow anaerobic sludge blanket, anaerobic fluidized bed reactor etc. This paper presents the preliminary steps undertaken in order to develop the prediction model using BPN for an anaerobic WasteWater Treatment System (WWTS) in terms of COD level. Data preprocessing techniques are carried out using various normalization methods. The purpose is to train the ANN using BPN algorithm with normalized input data and to evaluate the model performance with regard to MSE and Regression. The minimum MSE evaluates the accuracy and the Regression value, when obtained as 1, the difference between the actual and predicted values as minimum.

The Paper is further structured with various sections: Sect. 2 provides an overview of the background study, Sect 3 about the System Overview, Sect. 4 provides the Experimental results and Sect. 5 concludes the paper.

2 Related Works

2.1 Artificial Neural Network

Neural networks are used to solve problems in which the complete formulation is unknown i.e. no casual model or mathematical representations exist. The structure of ANN defines the overall architecture of the network, including one input layer, one output layer and usually one or two hidden layers. Each neuron receives a weighted sum from each neuron oh the next layer. Thus,

$$net = \sum_{i=1}^{n} W_i X_i \tag{1}$$

Where *net* is the summation of the input signal, and W_i, denotes an element of the weight vector W, and X_i is an element of the input vector X. For a given network and input vector, the output vector is totally determined by the weights. The process of finding optimal weight is to find optimal weight, called "training". The training algorithms used in this study were back-propagation. In this process, the input units and their desired output value were set for the network. The activations of the units were then calculated, feeding forward layer-by-layer from the inputs to the output.

2.2 Waste Water Treatment

UAF were fed with concentrated cheese dairy wastewaters. This reactor was operated for a total period of 197 days. UAF was initially fed at a low OLR of 0.4 g COD/L.d and an extended HRT of 77 d. OLR was then progressively increased in steps by 20–30 % once or twice a week provided that CODs removal efficiency remained above 80 %. A CODs removal efficiency of 80 % was considered as the threshold level for the operation of the reactor as described by Thanial et al. [6]. Efforts were made to maintain constant influent COD concentration, while the OLR was gradually increased by decreasing the hydraulic retention time (HRT). Maximum OLR corresponding to 80 % COD removal was thus ascertained for each reactor.

2.3 Reactor Study

The fixed bed reactor or anaerobic filter is a bioprocess in which the bacteria colonize in and around certain materials, which increase the usable surface area for the bacterial growth [6]. This kind of reactor allows a higher volume load and is suitable for wastewater treatment of high organic material content. The UAF is a double-walled laboratory-scale reactor at different operating conditions with an effective volume of 10 L and was equipped with a hot water jacket to maintain a mesophilic temperature of 33 ± 1 °C. The carrier material used was small buoying polyethylene packing media, which are cylindrical in shape (29 mm high and 30/35 mm diameter) and baffled with 16 compartments. Reactors were filled with these media entities to a height of about 75 cm (about 80 % of working volume of the reactor). The density and specific area of the media were 0.93 and

320 m^2/m^3 respectively. The reactor was fed in an upflow mode with concentrated chees-dairy wastewater with concentration of about 30 g CODt/L. The reactor was equipped with a continuous internal recirculation system from top to bottom of the reactor at the rate of 10 L/h. Recirculation was done mainly to eliminate the possibility of high organic loading close to the feed port and to achieve better contact of wastewater and sludge. The reactor was charged with anaerobic sludge (10 % by volume) collected from a methanogenic digester treating distillery vinasse. The concentration of volatile suspended solids (VSS) in the sludge was estimated to be around 21 g/L.

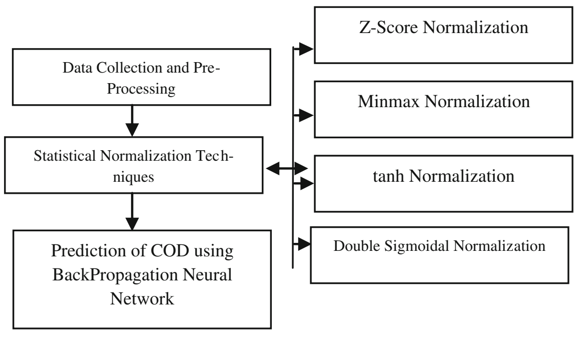
3 System Study

In this paper, cheese-dairy wastewater is considered and the wastewater treatment system includes the several phases as given in Fig. 1.

3.1 Data Collection and Pre-Processing

The task of processing the data is to establish the validity and quality of data from the system involving endless sources of disturbances. The sources contain electromagnetic interference, defective installation, insufficient maintenance or erroneous use and handles measuring system. The anaerobic filter treating agro-food wastewater during a period of 6 months with a total number of 200 samples is considered for the study. Most real-world databases contain outliers or missing, unknown, or erroneous data. Large number of missing values and outliers are accommodated in the multidimensional historical database of an UAF. In the anaerobic data, the total samples based on the effluent Chemical Oxygen Demand (COD) when using COD, Volatile Suspended Solid, Total Suspended Solid as inputs in the crude supply stream from the dataset is found to be treated up to 80 %.

Fig. 1 Overall process in anaerobic wastewater treatment system for the prediction of COD



Real data sets tend to be imperfect, contains errors, outliers, missing data, extra noise and tools either for detecting or correcting it are required. Data preprocessing is considered as prior analysis of data. Regression will work best if the number of WWTS parameters having missing values in their records is small. Outliers can sometimes be accommodated in the data through the use of trimmed means. In this section, brief outline of the K-Means clustering is given. The results of applying the algorithm is the removal of the missing values and outliers and then clustering is used to group the related data for processing [7].

3.1.1 K-Means Clustering

In this analysis the user starts with a collection of samples and attempts to group them into 'k' Number of Clusters based on certain specific distance measurements [8]. The general objective is to obtain that partition which, for a fixed number of clusters, minimizes the square-error. K-Means is a commonly used partitional algorithm with square-error criterion, which minimizes the following error function.

$$E = \sum_{k=1}^{C} \sum_{x \varepsilon Q_k} ||x_i - c_k||^2 \tag{2}$$

where C is the number of clusters, c_k is the center of cluster k and x is a data sample that belongs to cluster Q_k.

3.1.2 Statistical Normalization Algorithms

The statistical normalization techniques are used to pre-process the data. In this section all the normalization techniques are used individually to find the suitable normalization technique to improve the prediction accuracy of Back Propagation Neural network model in calculating the COD level.

Z-Score Normalization

The z-score computed the arithmetic mean and standard deviation of the particular data. Then it is necessary to calculate the mean and standard deviation of the scores approximately. The normalized scores are given by

$$s'_k = \frac{s_k - \mu}{\sigma} \tag{3}$$

where μ represents the arithmetic mean and σ represents the standard deviation of the given data. In case of an arbitrary distribution, mean and standard deviation are adequate estimates of location and scale, correspondingly, but are not finest.

If the resultant Z-score is greater than the corresponding criterion value based on the number of data points, then the suspect value is indeed an outlier, and should be excluded and the control limits re-calculated. The Z-Score normalization method is implemented and based on the Z-Score value, the outliers in the data set are removed.

Minmax Normalization

This technique rescales the features or outputs from one range to another new range of values. Generally, the features are rescaled within a specific range of 0–1 or from -1 to 1. The rescaling is typically accomplished by using a linear interpretation formula such as,

$$x' = (x_{\max} - x_{\min}) \times \frac{(x_i - x_{\min})}{(x_{\max} - x_{\min})} + x_{\min} \tag{4}$$

in which $(x_{\max} - x_{\min}) = 0$ when $(x_{\max} - x_{\min}) = 0$, for a feature, it specifies a stable value for that feature in the data. When the min–max normalization is implemented, each feature which fits within the new range of values will be stable.

tanh Normalization

The tanh-estimators are also robust and extremely effective [9]. The normalization is given by,

$$s'_k = 1/2\left\{\tanh\left(0.01\left(\frac{s_k - \mu GH}{\sigma GH}\right)\right) + 1\right. \tag{5}$$

where μ and σ represents the mean and standard deviation estimates, correspondingly, of the valid score distribution as specified by Hampel estimators. Hampel estimators depend on the following influence ()-function:

$$\psi(u) = \begin{cases} \{u, 0 \leq |u| < a \\ a\sin(u), a \leq |u| < b \\ a\sin(u)\left(\frac{c-|u|}{c-b}\right), b \leq |u| < c \\ 0, |u| \geq c \end{cases} \tag{6}$$

As a result, this technique is not responsive to outliers. If the effect of a huge number of tail-points is decreased, the effect is more robust but not effective. Alternatively, when many tail-points affect the approximation, the approximation is not robust but there must be considerable increase in the effectiveness. As a result, the parameters $a, b\, and\, c$ must be selected with attention based on the amount of robustness essential which in sequence depends on the estimate of the amount of noise exist in the available training data.

Double Sigmoidal Normalization

Double sigmoid function for normalization [10] is given by,

$$s'_k = \frac{\frac{1}{1+\exp(-z(s_{k-t}/r_1))} \, if s_k < t}{\frac{1}{1+\exp(-z(s_{k-t}/r_2))} \, otherwise} \tag{7}$$

where t represents the reference operating point and r_1 and r_2 indicates the left and right normalization points, i.e., the double sigmoid function reveals linear characteristics in the interval $(t - r_1, t - r_2)$ This technique transforms the scores into the [0,1] interval. However, it needs cautious tuning of the parameters t, r_1, r_2 to achieve better efficiency. The double sigmoid normalization is comparably related to the min–max normalization followed by the application of two-quadrics (QQ) or logistic (LG) function [10].

3.2 *Prediction of COD Using Back Propagation Neural Network*

Back propagation is the training technique usually used for this purpose. Levenberg–Marquardt algorithm is used to train the Back Propagation.

The performance of ANN model for Levenberg–Marquardt algorithm is evaluated by calculating the MSE as in (8) and Regression Coefficient (R) as in (9).

$$\mathrm{MSE} = \sum_{i=1}^{N} \frac{(x_i - y_i)^2}{N} \tag{8}$$

$$\mathrm{R} = \frac{\sum (y_i - \overline{y_i})}{\sqrt{\sum y_i^2 - \overline{y_i^2}}} \tag{9}$$

Thus the model was trained using Levenberg–Marquardt training algorithm and was found that the network model learnt to adjust the weights to predict the target outputs. If the level of learning is too large, then the algorithm will become unstable while on the contrary, algorithms take a long time to converge.

4 Experimental Results

Daily records from the operation of an anaerobic filter treating agro-food wastewater during a period of 6 months with a total number of 200 samples were obtained [11]. The raw data related to the anaerobic digestor and each sample comprises the parameters that are non linear. In the anaerobic data, the total samples is found to be treated up to 80 % based on the effluent Chemical Oxygen Demand (COD), when considering COD, Volatile Suspended Solid, Total Suspended Solid as inputs in the crude supply stream. The influent COD levels, the controlling parameters of anaerobic reactor were considered as the input parameters. The model was trained to recognize these input parameters in order to predict the corresponding effluent level of COD as output parameters. The model was trained with the data in the training subset with influent COD levels, flow rate and OLR to optimize the network weights so as to minimize the appropriate error function. After preprocessing, the essential features being influent COD(t), influent COD(s), flow rate, volatile suspended solids and total suspended solids are selected and given as inputs .

In this paper, the data was analysed by applying K-Means algorithm in Java. The number of records processed is 200 and the processing time to recognize and eliminate the missing values and outliers was found to be minimum expressed in nanoseconds. Thus the missing values and outliers were recognized and 61 records were eliminated with the processing time of 78 ns. The K-Means partitional clustering algorithm is computationally efficient and provides good results if the clusters are compact. The prediction of COD by BPN based on accuracy and execution time is improved by various statistical normalization techniques. The experimental results show that the performance of the prediction model using the neural networks is dependent on the normalization methods. The z-score normalization method is found to be the suitable one to normalize the data than other existing normalization techniques. The data set prepared using normalization enhance the learning capability of the network with minimum error. Based on the experimental results, Back propagation approach is found to be best for the prediction of COD.

4.1 Prediction Accuracy

Prediction accuracy is evaluated for normalization techniques in the prediction of COD level. In order to evaluate the statistical normalization techniques for the prediction of COD in the back propagation neural network model, prediction accuracy is calculated and the results are given in Fig. 2.

4.2 Execution Time

The execution time is calculated based on the machine time (i.e., the time taken by the machine to run the proposed algorithm) for predicting the COD level and it is provided in Fig. 3.

The main purpose of this paper is to find the suitable statistical normalization techniques for the prediction of COD level and to improve the prediction accuracy

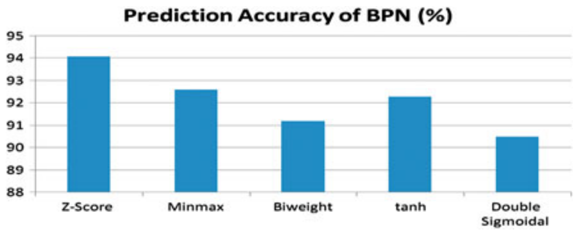


Fig. 2 Comparison of prediction accuracy for the normalization techniques in predicting the COD

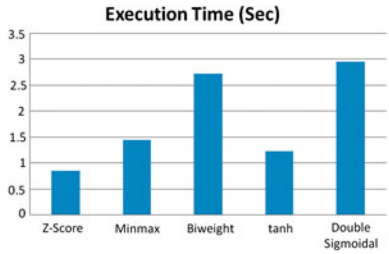


Fig. 3 Comparison of time taken the normalization techniques in predicting the COD

after removing the noisy data. The model was developed by using Matlab software and the performance of Back propagation model for each normalization techniques was evaluated by calculating the prediction accuracy and execution time.

4.3 MSE and Regression of LM Algorithm

The data normalized by z-score method is then trained using LM algorithm. The trained model is tested with the test data subset and the model was able to recognize the inputs and the model performance was found satisfactory based on the MSE and R. The performance of the network model is evaluated by computing the Regression and MSE for each trial made. The LM training algorithm had its impact on the model for training with minimum errors and the goal being met with the minimum number of epochs. The minimum MSE obtained for the LM algorithm showed the accuracy of the trained network in recognizing the input and predicting the corresponding output as shown in Fig. 4.

The R value gained shows the best fit of the network model in the prediction of the corresponding output close to the measured data. The MSE value for LM algorithm is 0.53375. Figure 5 shows the R value obtained for LM algorithm for the training, validation and testing phases as 0.99185, 0.8108 and 0.8145 respectively.

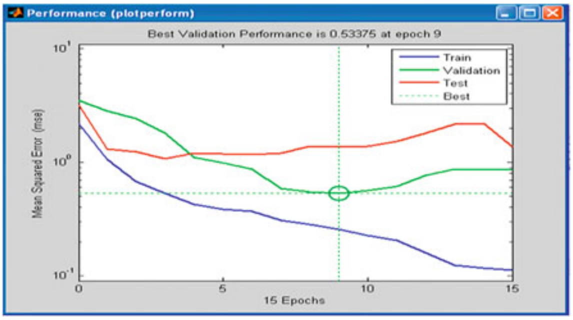


Fig. 4 Performance graph with MSE for Levenberg–Marquardt algorithm

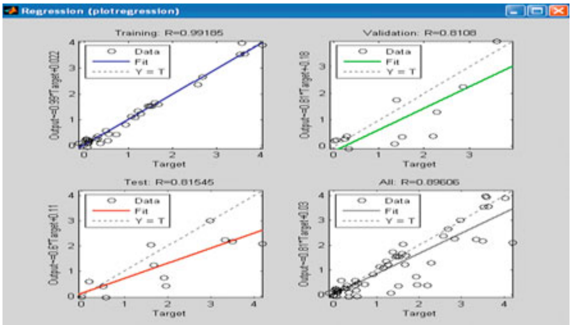


Fig. 5 Regression graph for Levenberg–Marquardt algorithm

5 Conclusion

The significance of data preparation for a BPN model developed using Levenberg–Marquardt (LM) algorithm is presented in this paper. Input data pretreatment performed based on k-means clustering provided the quality data. Further data is scaled to fall within a small, specified range by z-score normalization. The BPN provided good estimates of COD levels covering a range of data for training and testing purposes for the normalized data. The performance of the network was found satisfactory based on Regression and MSE values 0.99185 and 0.53375 respectively.

References

1. Hamed MM, Khalafallah MG, Hassanien EA (2004) Prediction of wastewater treatment plant performance using artificial neural networks. Environ Model Softw 19:919–928
2. Plazl I, Pipus, Grolka M, Koloini T (1999) Parametric sensitivity and evaluation of a dynamic model for single stage wastewater treatment plant. Acta Chimica Slovenica 46(2):289–300

3. Hamoda MF, Al-Gusain IA, Hassan AH (1999) Integrated wastewater treatment plant performance evaluation using artificial neural network. Water Sci Technol 40:55–69
4. Dixon M, Gallop JR, Lambert SC, Healy JV (2005) Experience with data mining for the anaerobic wastewater treatment process. Environ Model Softw 22:315–322
5. Bernard O, Hadj-Sadok Z, Dochain D, Genovesi A, Steyer JP (2001) Dynamical model development and parameter identification for an anaerobic wastewater treatment process. Biotechnol Bioeng 75(4):424–438
6. Thanial JV, Torrijos M, Habouzit F, Moletta R (2007) Treatment of distillery vinasse in a high rate anaerobic reactor using low density polyethylene supports. Water Sci Tech 56(2):17–24
7. Barnett V, Lewis T (1994) Outliers in statistical data, 3rd edn. Wiley, Chichester
8. Rustum R, Adeloye AJ (2007) Replacing outliers and missing values from activated sludge data using Kohonen self-organizing map. ASCE 133(9):909–916
9. Qingling G (2011) A generalized tanh method and its application. Appl Math Sci 5(76):3789–3800
10. Modugno R, Pirlo G, Impedovo D (2010) Score normalization by dynamic time warping. In: IEEE international conference on computational intelligence for measurement systems and applications (CIMSA), pp 82–85
11. Vijayabhanu R, Radha V (2010) Recognition and elimination of missing values and outliers from an anaerobic wastewater treatment system using K-Means cluster. In: 3rd international conference on advanced computer theory and engineering (ICACTE), vol 4. pp 186–190
12. Rajinikanth R, Ganesh R, Escudie R, Mehrotra I, Kumar P, Thanikal JV, Torrijos M (2009) High rate anaerobic filter with floating supports for the treatment of effluents from small-scale agro-food industries. Desalin Water Treat 4:183–190

Contrast Improvement of Mammographic Masses Using Adaptive Volterra Filter

Ashutosh Pandey, Anurag Yadav and Vikrant Bhateja

Abstract Due to ill-performance of X-ray hardware systems, mammographic images are generally noisy with poor radiographic resolution. This leads to improper visualization of lesion details. This paper presents an improved Volterra filter design known as Adaptive Volterra filter for contrast enhancement of mammograms. The operation of the adaptive filter proposed in this work can be classified as Type-0, Type-1 and Type-2 depending upon the nature of background tissues (fatty, fatty-glandular or dense) in the mammogram. This filter is considered as a Taylor series with memory whose truncation to the first non-linear term may lead to a simpler and effective representation. Computer simulations are performed on digital mammograms from MIAS database yielding promising improvement in contrast of the targeted lesion along with reasonable suppression of background in comparison to other enhancement techniques.

Keywords Breast cancer · Contrast improvement index · MIAS database · Quadratic filter · Volterra filter

A. Pandey (✉) · A. Yadav · V. Bhateja
Department of Electronics and Communication Engineering, Shri Ramswaroop Memorial Group of Professional Colleges, Faizabad Road, Lucknow 227105 UP, India
e-mail: ashutosh91.p@gmail.com

A. Yadav
e-mail: anurag.yadav.ece@gmail.com

V. Bhateja
e-mail: bhateja.vikrant@gmail.com

Mohan S. and S. S. Kumar (eds.), *Proceedings of the Fourth International Conference on Signal and Image Processing 2012 (ICSIP 2012)*, Lecture Notes in Electrical Engineering 222,
DOI: 10.1007/978-81-322-1000-9_54,

1 Introduction

American Cancer Society reported Breast cancer as the second largest cause of death among the women with nearly 39,510 deaths estimated in 2012 in America only. According to National Cancer Institute breast cancer develops in the tissues of breast, usually the ducts and lobules and becomes invasive, once it reaches the healthy tissues. Early detection of breast cancer is an effective way to reduce mortality, once detected in early stages, it can be cured successfully. Mammography is the most effective, reliable and cheapest technique adapted by radiologists for the initial screening and diagnosis of breast cancer. But due to limitations of X-ray hardware systems the screened mammograms experiences poor radiological resolution. Due to this the benign and the malignant tissue is not readily discernable and true diagnosis of breast cancer get impeded [1]. It is over here that the role of mammogram enhancement techniques comes into process for contrast improvement of the targeted lesion with respect to its background. Some of the mammogram enhancement technique developed so far are reviewed and discussed in the subsequent paragraphs. Lojacono et al. [2] developed an algorithm for mammogram enhancement using combination of Dyadic wavelets and morphological techniques. This algorithm was adaptive in nature for different types of anomalies but the improvement in contrast to clearly differentiate between healthy tissues and tumor was only marginal. Also, the authors have used separate algorithms for denoising, enhancement and segmentation. Mammogram enhancement techniques employing fuzzy domain transformation [3] satisfactorily removed the adverse effects of noise but were less flexible in approach, as more than one parameter has to be varied for different types of images. On the other hand, enhancement using adaptive histogram equalization [4] was limited to be applied on mammograms affected by low noise density. Chang and Laine used coherence of multiscale features [5] for enhancement of mammograms. However, the results depicted very low contrast enhancement and the practical implementation of this algorithm was also complex due to its multistage operation. Yan et al. proposed approximation weighted detail contrast enhancement (AWDCE) [6] for detection of masses in mammogram using Daub20 wavelet Transform. The results showed improvement in contrast of masses but contrast enhancement index compared was close to unity along with considerable erosion of mass. Zheng et al. [7] used skeletonization and mesh formation to increase the resolution of the mammograms, but the fine details of the tumors were not visible. Al-Kindi and Al-Kindi [8] utilized a combination of histogram equalization and Cannys' edge detection to enhance the contrast of sonograms and mammograms. However, the edges and boundaries of the tumor in the obtained results were eroded. The remaining part of paper is organized as follows: Sect. 2 gives the background of non-linear filters along with description of proposed filter. Section 3 details the obtained simulation results and their analysis. Conclusions on the basis of the obtained results are given under Sect. 4.

2 Proposed Filter Design Methodology

2.1 Background

Linear filters [9] generally exhibit simplicity in design, analysis and synthesis, but do not give impressive results for images coupled with signal-dependent or multiplicative noises as well as for those with Non-Gaussian statistics. Noises removed by the linear filters often leads to image blurring as edges could not be preserved. To overcome this drawback non-linear filters came into existence [9, 10]. These filters are created by models which utilize Volterra filters [9], order statistics filters [11], and morphological filters [12]. The use of non-linear filters provide better image filtering results not only by suppressing effect of noises but also by preserving edges of the image. The only demerit possessed by non-linear filters is that it requires a large number of coefficients for designing. Hence, many techniques are developed to reduce the number of independent weighted coefficients of the non-linear filters [13, 14]. The quadratic filters are the simplest form of the non-linear filters. But, do not possess simpler characterization as required for linear filters in frequency domain [15]. Volterra series [9] offers a simplified and manageable approach to introduce non-linear effects during conventional linear analysis. Volterra filter is a type of non-recursive filter as their output is dependent only on the present and past values of input samples thereby leading to simpler and effective representation without having prior knowledge of higher order statistics. These filters find significant usage in various image processing applications like: contrast enhancement, edge detection, segmentation and also in restoration of images blurred by physical phenomena.

This paper presents a new Adaptive Volterra Filter (AVF) for enhancement of digital mammograms. The operational response of the proposed filter is categorized into three different types on the basis of the category of background tissues in the mammograms. These filters provides promising contrast enhancement of lesion with due suppression of background noises.

2.2 Generalized Form of Adaptive Volterra Filter

A novel class of Adaptive Volterra filters (AVF) is introduced in this work for enhancement of digital mammograms. It is a discrete time invariant non-linear filter equipped with memory, and can be expressed as a discrete Volterra series expansion.

$$y(n) = \beta_0 + \sum_{L=1}^{\infty} \beta_L[x(n)] \tag{1}$$

where: *y(n)* and *x(n)* are the output and input images respectively; *n* represents the pixel gray level value at a particular location *(i, j)*. β_L represents the *L*th-order AVF and the constant β_0 denotes an offset term.

AVF proposed in this work can be generated by second order truncation on the upper limit of Volterra series (defined in (1) above) to yield the following input–output relationship:

$$y(n) = \underbrace{\sum_i \theta(i) x^{2\gamma(i)}(n-i)}_{y_{linear}} + \underbrace{\sum_i \sum_j \phi(i,j) x^{\lambda(i)}(n-i) x^{\lambda(j)}(n-j)}_{y_{quadratic}} \tag{2}$$

where: $\theta(i)$ and $\phi(i,j)$ represent the linear and quadratic filter coefficients. γ is the weight index for the linear term in (2). In this case, the input pixels of the linear filter are raised to the power of 2γ. Similarly, each of the collected and grouped input pixels for the quadratic filter in (2) is raised to the power of λ.

2.3 Determination of Filter Coefficients for AVF Realization

A 3×3 AVF filter used in this work comprises of linear filter coefficients represented as a 3×3 matrix and the quadratic filter coefficients represented with a 9×9 matrix. Realization of AVF involves computation of a linear component consisting of 9 coefficients and quadratic component of 81 independent coefficients. This design complexity is simplified by utilizing the properties of symmetric and isotropic kernels. In order to preserve an untransformed pixel at the output the sum of linear coefficients must be made equal to unity and that of quadratic coefficients to zero.

$$\sum_i \theta(i) = 1 \tag{3}$$

$$\sum_i \sum_j \phi(i,j) = 0 \tag{4}$$

The filter operator must be isotropic in order to produce filter response independent of features like edges or textures of the input image.

$$\theta(i) = \theta(N-1-i),\ i = 0,\ 1, \ldots, N-1 \tag{5}$$

$$\phi(i,j) = \phi(N-1-i,\ N-1-j),\ i, j = 0, 1, \ldots, N-1 \tag{6}$$

By utilizing kernel symmetry, the number of 81 independent coefficients of the quadratic component gets reduced to 45 in which each of them is involved in one of the 13 independent responses. By applying isotropic property, only 11 independent coefficients and 6 impulse responses are left. For AVF realization, a kernel of size 3×3 is used in this work, for which (3) reduces to:

$$4\theta_1 + 4\theta_2 + \theta_0 = 1 \tag{7}$$

Optimal selection of values of the filter coefficients θ_0, θ_1 and θ_2 in (7) will be made in such a manner that they satisfy the above equation. Similarly, values of 11 independent coefficients of the quadratic filter (ϕ_0–ϕ_{10}) will be optimally selected to satisfy Eq. (8) [obtained from (4) for a 3 × 3 kernel].

$$4\phi_1 + 16\phi_3 + 8\phi_7 + 8\phi_4 + 16\phi_{10} + 4\phi_8 + 4\phi_2 + 8\phi_6 + 8\phi_5 + 4\phi_9 + \phi_0 = 0 \tag{8}$$

2.4 Classification of AVF

With the determination of filter coefficients for both linear and quadratic components using a 3 × 3 kernel the generalized version of the proposed AVF defined in (2) can be expressed in the following form:

$$y(n) = y_{linear} + y_{quadratic} \tag{9}$$

The linear and quadratic components of (9) can be further classified into three different categories as; Type-0, Type-1 and Type-2 based upon the distance between two pixels of a 3 × 3 kernel of the input image as explained in Fig. 1.

Type-0 AVF consists of input image pixels in which the difference between the two pixels is zero in a 3 × 3 filter kernel (refer Fig. 3a). Hence for Type-0 AVF, the linear and quadratic components of (9) can be defined as:

$$y_{linear} = \theta_0 x_5^{2a} + \theta_1(x_1^{2b} + x_3^{2b} + x_7^{2b} + x_9^{2b}) + \theta_2(x_2^{2c} + x_4^{2c} + x_6^{2c} + x_8^{2c}) \tag{10}$$

$$y_{quadratic} = \phi_0 x_5^{2a} + \phi_1(x_1^{2b} + x_3^{2b} + x_7^{2b} + x_9^{2b}) + \phi_2(x_2^{2c} + x_4^{2c} + x_6^{2c} + x_8^{2c}) \tag{11}$$

Similarly, in a 3 × 3 kernel of *Type-1 AVF,* the difference between the two input pixels is one; namely the two image pixels are adjacent to each other. For this type of filter, (9) is defined as:

$$y_{linear} = \theta_0 x_5^{2a} + \theta_1(x_1^{2b} + x_3^{2b} + x_7^{2b} + x_9^{2b}) + \theta_2(x_2^{2c} + x_4^{2c} + x_6^{2c} + x_8^{2c}) \tag{12}$$

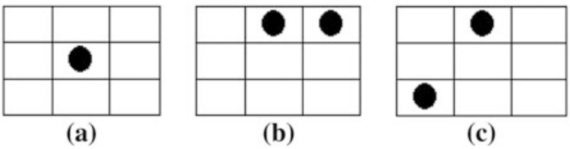


Fig. 1 Distance between the pixels locations in a 3 × 3 kernel used for categorizing AVF. **a** Type-0 AVF (both pixels on same location), **b** type-1 AVF (pixels on different location having distance of one unit) and **c** type-2 AVF (pixels having distance of two units)

$$\begin{aligned} y_{quadratic} &= \phi_3(x_1^b x_2^c + x_1^b x_4^c + x_2^c x_3^b + x_3^b x_6^c + x_4^c x_7^b + x_6^c x_9^b + x_7^b x_8^c + x_8^c x_9^b) \\ &+ \phi_4(x_1^b x_5^a + x_3^b x_5^a + x_5^a x_7^b + x_5^a x_9^b) + \phi_5(x_2^c x_5^a + x_4^c x_5^a + x_5^a x_6^c + x_5^a x_8^c) \\ &+ \phi_6(x_2^c x_4^c + x_2^c x_6^c + x_4^c x_8^c + x_6^c x_8^c) \end{aligned} \tag{13}$$

Type-2 AVF comprises the input image pixels in which the difference of two pixels is two in a 3 × 3 filter kernel. For this type of filter, (9) can be stated as:

$$y_{linear} = \theta_0 x_5^{2a} + \theta 1\left(x_1^{2b} + x_3^{2b} + x_7^{2b} + x_9^{2b}\right) + \theta_2\left(x_2^{2c} + x_4^{2c} + x_6^{2c} + x_8^{2c}\right) \tag{14}$$

$$\begin{aligned} y_{quadratic} &= \phi_7(x_1^b x_3^b + x_1^b x_7^b + x_3^b x_9^b + x_7^b x_9^b) + \phi_8(x_1^b x_9^b + x_3^b x_7^b) \\ &+ \phi_9(x_2^c x_8^c + x_4^c x_6^c) + \phi_{10}(x_1^b x_6^c + x_1^b x_8^c + x_2^c x_7^b + x_2^c x_9^b + x_3^b x_4^c \\ &+ x_3^b x_8^c + x_4^c x_9^b + x_6^c x_7^b) \end{aligned} \tag{15}$$

Imposition of condition of distance between two image pixels for framing the quadratic component (11), (13) and (15) would lead to cancelation of input pixels existing in isolation or in adjacent pairs. This will have no effect on pair of pixels having inter-pixel distance of zero, one and two units (in Type-0, Type-1 and Type-2 AVF respectively). This filter would be therefore capable in discrimination of texture patterns formed by adjacent or isolated pixels. The categorization of filter design highlights the adaptive nature of the filter as experimentally each of these types will be applied for a specific category of mammograms. The weight indices γ and λ of (2) are substituted with *a, b* and *c* as shown in (10)–(15) above. In a 3 × 3 kernel *a, b* and *c* are the powers on the pixels indicated at the shaded locations shown in Fig. 2a–c respectively. The values for these powers will be determined experimentally and are optimally tuned depending upon the values of performance evaluation parameters (*CII*). The above design constraints leads to development of filter suitable for enhancement coupled with significant suppression of background noise.

3 Results and Discussions

3.1 Experimental Procedure

The digital mammograms used in this work for simulation are taken from the Mammographic Image Analysis Society (MIAS) database [16]. MIAS is a UK based organization involved in mammogram related researches, and has generated

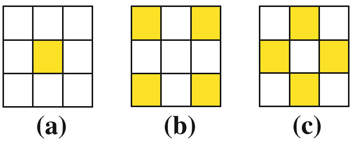


Fig. 2 Pixel locations in a 3 × 3 kernel. **a** Centre pixel, **b** pixels at odd numbered locations and **c** pixels at even numbered locations

a database of 322 digital mammograms. It is the easiest available mammogram database for the research work. Simulations are performed on nearly 40 mammograms containing masses embedded in various types of background tissues, i.e. fatty-glandular, fatty and dense glandular. Digital mammograms (of size 1024×1024) obtained from the database are initially normalized before further processing. In these test images, the region containing the tumor is treated as foreground and the remaining area as background. As explained under Sect. 2.3, for a 3×3 kernel, the values of the linear and quadratic filter coefficients are experimentally determined. These are those set of coefficient which satisfy the Eqs. (7) and (8). The performance evaluation of an enhancement filter for mammograms can be ascertained by its capability to improve the difference between the mean gray levels lying in the image foreground with respect to the background. The objective evaluation of the proposed filter and its performance comparison with other mammogram enhancement techniques has been made using Contrast Improvement Index (*CII*), Peak Signal-to-Noise Ratio (*PSNR*) and Average Signal-to-Noise Ratio (*ASNR*) [17]. Higher the values of *CII, PSNR,* and *ASNR* more promising is contrast enhancement filter. The experimentally determined coefficients values are optimally selected using *CII* as the performance parameter. In this work, the values of linear and quadratic filter coefficients experimentally determined for implementation are: $\theta_0 = 0.2$, $\theta_1 = \theta_2 = 0.1$, $\phi_0 = 8€$, $\phi_1 = \phi_2 = \phi_6 = -€$, $\phi_3 = -0.5€$, $\phi_4 = \phi_5 = \phi_{10} = €$, $\phi_7 = -2€$, $\phi_8 = -4€$, $\phi_9 = 4€$. Significantly improved enhancement results are obtained for $€ = 0.1$. In a similar fashion, the values of the weight indices are taken as: $a = 8\mu$, $b = c = \mu$; where μ has different range of values for the three different types of proposed AVF. For Type-0, Type-1 and Type-2 AVF; μ varies between 0.3–0.5, 0.5–0.7 and 2.8–3.4 respectively.

3.2 Simulation Results

Once the optimal values of filter coefficients as well as weight indices are determined, the input test mammograms are then processed with the proposed AVF given in (9)–(15). The adaptive behavior of the proposed filter is based on the distinguishing filter response depending upon the nature of background tissues (fatty-glandular, fatty and dense-glandular). From the enhancement results given in Fig. 3, it can be visualized that the proposed Type-0 AVF not only enhances the targeted tumor region in the mammogram having fatty glandular background tissues but also preserves the finer details. Similarly, the proposed Type-1 and Type-2 AVF gives much better enhancement results on the mammogram with fatty and dense glandular background tissues respectively. For the sake of comparison, the enhancement of the test mammograms is also performed using conventional enhancement techniques like: Unsharp Masking (UM) [18] and CLAHE [19] as well as recently developed enhancement approaches, which includes: Quadratic

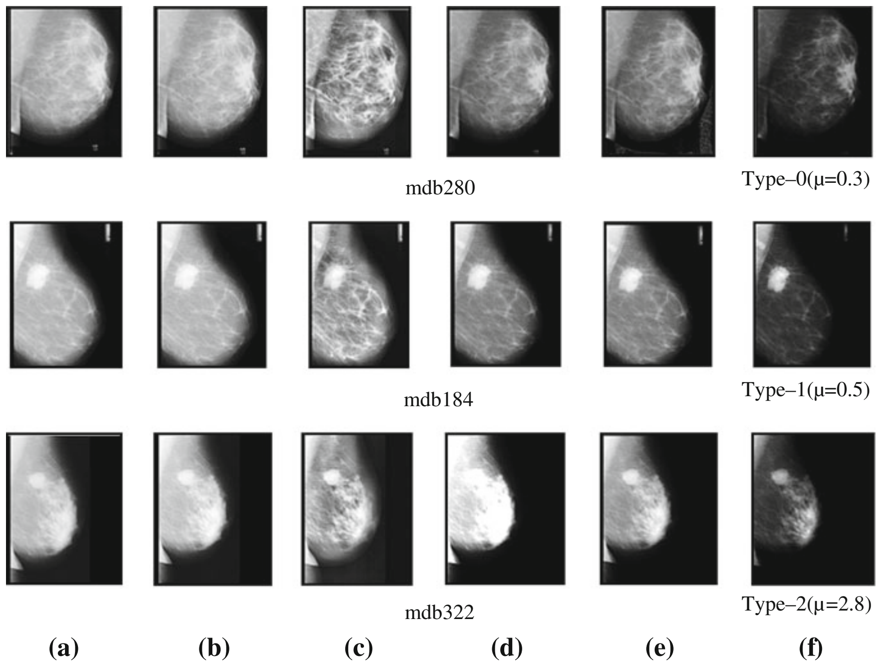


Fig. 3 **a** Original mammograms. Mammograms processed with different enhancement techniques. **b** UM [18]. **c** CLAHE [19]. **d** QF [20]. **e** AWQF [21]. **f** Proposed AVF

Filter (QF) [20] and Alpha Weighted Quadratic Filter (AWQF) [21]. The original as well as the enhanced mammograms for the three test images are shown in Fig. 3a–f.

In the results obtained by other mammogram enhancement techniques, the white tumor region is camouflaged with the background tissues. These techniques are not able to provide reasonable suppression of the background. The values of *CII, PSNR* and *ASNR* are computed for the enhanced images and are tabulated under Tables 1, 2, 3.

3.3 Comparison

Higher values of *CII* depict a higher degree of enhancement obtained by an enhancement technique. From the *CII* results given under Tables 1, 2, 3, it can be ascertained that QF [20] and AWQF [21] enhances contrast of the tumor region better than the images enhanced using UM [18] and CLAHE [19]; but still it fails to suppress the background effectively. In the case of UM [18], CLAHE [19], QF [20] and AQWF [21] lower values of *PSNR* and *ASNR* clearly explains that the noise levels in the transformed images are not reduced. The tabulated results

Table 1 Performance comparision of different enhacement techniques on mammogram (mdb280) with fatty-glandular background tissues

	CII	PSNR	ASNR
UM [18]	1.180264	1.039060	1.187852
CLAHE [19]	1.074817	1.303545	1.186112
QF [20]	1.828134	1.438666	1.724130
AWQF [21]	1.841925	1.464655	1.743199
AVF (Type-0)	2.288806	2.706213	1.984083
AVF (Type-1)	1.960530	2.013452	1.816248
AVF (Type-2)	2.172316	2.305842	1.875241

Table 2 Performance comparision of different enhacement techniques on mammogram (mdb184) with fatty background tissues

	CII	PSNR	ASNR
UM [18]	1.019045	0.998189	0.953638
CLAHE [19]	0.783021	0.844278	0.600212
QF [20]	1.779677	1.150828	0.986434
AWQF [21]	1.772233	1.229548	0.977323
AVF (Type - 0)	2.809334	2.197819	1.566872
AVF (Type - 1)	3.215213	2.425499	1.636591
AVF (Type - 2)	3.045453	2.317468	1.595417

Table 3 Performance comparision of different enhacement techniques on mammogram (mdb322) with dense-glandular background tissues

	CII	PSNR	ASNR
UM [18]	1.222540	1.085059	1.131288
CLAHE [19]	1.535115	1.451130	1.471133
QF [20]	2.450219	1.383008	2.252803
AWQF [21]	1.697883	1.419595	1.466825
AVF (Type - 0)	3.284144	2.139028	2.710270
AVF (Type - 1)	2.993076	2.131675	2.577032
AVF (Type - 2)	3.750198	3.114328	3.239086

clearly differentiate the performance of three different categories of proposed AVF. Results of Table 1 clearly depicts that the proposed Type-0 AVF gives much better enhancement results on the mammogram having fatty-glandular background tissue in comparison to other two types of AVF. This filter (Type-0) not only enhances the foreground but also suppress the background noises. It preserves both i.e. the edges of the lesion and finer details of the mammographic image. Similarly, the results of Tables 2 and 3 clearly depicts that the performance of proposed Type-1 and Type-2 AVF is much better on mammogram with fatty and dense-glandular background tissues respectively than other type of proposed AVF. Here, it can be said that different adaptive design approaches has been adopted for the three different proposed AVF in order to provide better enhancement results on mammograms with different background tissues. Finally the higher values of *CII, PSNR* and *ASNR* validate that the performance of three proposed AVFs is much better in comparison to other mammogram enhancement techniques.

4 Conclusion

Non-linear filtering techniques provide better contrast enhancement results not only by suppressing the background noises but also preserving subtle finer details of the targeted lesion. In this paper a novel design of Adaptive Volterra filter (AVF) is proposed for enhancement of digital mammograms. The three different types of the proposed AVF i.e. Type-0, Type-1 and Type-2 are adaptively designed to address the mammographic masses embedded in background of different types of breast tissues. Experimental results clearly demonstrate that the proposed filter yields promising results with higher *CII* and values of signal-to-noise ratio. This ensures significant contrast improvement along with noise filtering and preservation of morphological details, justifying the application of the proposed filter for the radiologists in early diagnosis of breast cancer.

References

1. American Cancer Society: Cancer Facts and Figures (2012). http://www.cancer.gov/cancertopics/types/commoncancers
2. Lojacono R et al (2008) Mammographic images enhancement and denoising for breast cancer detection using dyadic wavelet processing. IEEE Trans Instrumen Meas 57(7):1422–1430
3. Tang H, Zhuang T, Wu Ed X (2001) Realization of fast 2-D/3-D image filtering and enhancement. IEEE Trans Med Imaging 20(2):132–140
4. Jusman Y, Mat Isa NA (2010) A proposed system for edge mammogram image. In: Proceedings of the of the 9th world scientific and engineering academy and society (WSEAS) international conference on applications of electrical engineering, pp 117–123
5. Chang C, Laine A (1999) Coherence of multiscale features for enhancement of digital mammogram. IEEE Trans Inf Technol Biomed 3(5):32–46
6. Yan Z et al (2001) An approximation-weighted detail contrast enhancement filter for lesion detection on mammograms. In: Proceedings of the 23rd annual conference of the IEEE on Engineering in Medicine and Biology Society, vol 3, pp 2472–2475
7. Zheng J, Fuentes O, Leung M (2010) Super resolution of mammograms. In: IEEE symposium on computational intelligence in bioinformatics and computational biology (CIBCB), Montreal, pp 1–7
8. Al-Kindi SG, Al-Kindi GA (2011) Breast sonogram and mammogram enhancement using hybrid and repetitive smoothing-sharpening technique. In: Proceedings of the 1st middle east conference on biomedical engineering (MECBME), Sharjah, pp 446–449
9. Mathews VJ, Sicuranza GL (1993) Volterra and general volterra related filtering. In: IEEE winter workshop on nonlinear digital signal processing, pp T_2.1–T_2.8
10. Salmond DJ et al (1993) Novel approach to non-linear/non-Gaussian Bayesian state estimation. In: Proceedings of the IEEE on radar and signal processing, vol 140, pp 107–113

11. Pitas I, Venetsanopoulos AN (1992) Order statistics in digital image processing. Proc IEEE 80(12):1893–1921
12. Stevenson R, Arce G (1987) Morphological filters: statistics and further syntactic properties. IEEE Trans Circ Syst 34(11):1292–1305
13. Mathews VJ (1991) Adaptive polynomial filter. IEEE Signal Process Mag 8(3):10–26
14. Sicuranza GL (1992) Quadratic filters for signal processing. In: Proc IEEE 80(8):1263–1285
15. Sicuranza GL (2000) Volterra filters for image and video processing. In: Proceedings of the first international workshop on image and signal processing and analysis, Pula, pp 15–26
16. Suckling J et al (1994) The mammographic image analysis society mammogram database. In: Proceedings of 2nd international workshop on digital mammography, York, UK, pp 375–378
17. Morrow WM et al (1992) Region based contrast enhancement of mammograms. IEEE Trans Med Imaging 11:392–406
18. Rogowska J, Preston K, Shasin D (2009) Evaluation of digital unsharp masking and local contrast stretching as applied to chest radiology. IEEE Trans Inf Technol Biomed Eng 35(2):236–251
19. Pisano ED et al (1998) Contrast limited adaptive histogram equalization image processing to improve the detection of simulated spiculations in dense mammograms. J Digit Imaging 11:193–200
20. Ramponi G (1990) Bi-Impulse response design of isotropic quadratic filters. Proc IEEE 78(4):665–667
21. Zhou Y et al (2009) Mammogram enhancement using alpha weighted quadratic filter. In: Proceedings of annual international conference of the IEEE Engineering in Medicine and Biology Society, Minneapolis, Minnesota, pp 3681–3684

Image De-Noising by Enhanced Median Filtering for High Density Noisy Images

Vikas Gupta and Abhishek Sharma

Abstract In the field of digital image processing [1], noise removal is always a critical process. In this paper we proposed an enhanced method of image de-noising. The purpose of this new method is to improve the signal to noise ratio (SNR) of de-noised image and get more better image, especially when image corrupted by high noise density. We improved the median filter algorithm, and get comparatively better results than previous methods. The mathematical analysis shows that this process improve the PSNR [2](Peak signal to noise ratio) at high density noise level. It also reduces the complexity of calculation because noise detection and noise removal both processes are performing simultaneously. This method produce better image without blurring and also preserve the edge and fine details of image.

Keywords Median filter · Threshold · Peak signal to noise ratio · Mean square error · High density noise

1 Introduction

During the transmission and acquisition, true values of pixels are affected by different type of noise. These noise can be Gaussian type or impulse noise [1]. The most common noise which badly affects the image is impulse noise. Impulse noise is of two types

V. Gupta (✉) · A. Sharma
Department of Electronics and Communication, Technocrats institute of technology, Anand nagar, Bhopal, India
e-mail: vgup24@yahoo.com

A. Sharma
e-mail: abhishektit09@gmail.com

Mohan S. and S. S. Kumar (eds.), *Proceedings of the Fourth International Conference on Signal and Image Processing 2012 (ICSIP 2012)*, Lecture Notes in Electrical Engineering 222,
DOI: 10.1007/978-81-322-1000-9_55,

(1) Fixed value impulse noise and
(2) Random valued impulse noise.

Fixed valued impulse noise [3] is fixed type of noise which can be either 0 or 255 (0 for black and 255 for white). Random valued impulse noise can have any value between 0 and 255. So the different types of techniques are requiring removing different types of noise. Filters are best suited for image de-noising, Because the filters can be easily realise physically. There were so many filters developed for image de-noising. The mean filter or average is very simple and easy filter for image noise removal but while removing the noise, mean filter affects fine details of image and causes blurring into the image. Also the nature of impulse noise is non-linear so the non-linear filter will work most efficiently then mean filter.

Median filter also called simple median filter (SMF) [1] is most common non-linear filter. The base of median filter is sorting. First it chooses the filtering window of size n × n, where n is odd number. Then it sorts the all elements of the selected filtering window. Then it chooses the middle value of the sorted elements and then replaces this median value with the central noisy pixel. After that window will slides to the next pixel and repeat the process. This filter gives the better results than mean and other linear filters. But it also has drawback that this filter works uniformly over the entire image. That causes blurring into the image. That is why some improvement has been made in median filter to get better results. Weighted median filter (WMF) is an improvement in the median filter. Some other filters are centre weighted median filter (CWM) [4], Adaptive median filter (AMF), Adaptive centre weighted median filter (ACWM) [5, 6], Switching median filter (SMF) [7], progressive switching median filter (PSMF) [8]. The combination of median filter with some other filter is also use to make improvement in the performance of median filter such as median filter combine with average filtering and median filter with high pass filter.

2 Noise Modal

Noise may be modelled as impulse noise or salt and pepper noise. The pixels corrupted by any of the fixed valued impulse noise (0 or 255). The corrupted pixels take either 0 (black) or 255 (white) with equal probability distribution.

$$g(x) = \begin{cases} P/2 & for \cdots x = 0 \\ 1 - P & for \cdots x = S(i,j) \\ P/2 & for \cdots x = 255 \end{cases} \tag{1}$$

Here P = Noise density,
g(x) = probability density function,
S(i,j) = intensity value,
X(i,j) = Noisy pixel.

3 Related Work

Median filter [9] is a simple non-linear filter for image de-noising. In this filter, we remove noise from the image by replacing the targeted noisy pixels by median value of its neighbours. The number of neighbours depends upon the size of filtering window. The median value is the value of middle element in a sorted sequence of filtering window.

$$\begin{aligned}\mathrm{Median(Z)} &= \mathrm{Med}\ \{\mathrm{Zi}\} \\ &= \begin{cases} \mathrm{Zi(n+1)/2}, & \text{n is odd} \\ 1/2[\mathrm{Zi(n/2)} + \mathrm{Zi(n/2)} + 1], & \text{n is even} \end{cases}\end{aligned} \tag{2}$$

Z1, Z2, Z3 ....Zn is the sequence of neighbour pixels. To apply median filter algorithm, all pixels need to be sorted in ascending or descending order. after sorting these pixels, the sequence will be Zi1 $\leq$ Zi2 $\leq$ Zi3 $\leq$ ....Zin, N is generally odd.

The size of filtering window [9] is an important factor. Smaller window size has good image details preservation but poor noise removal quality, and larger window size remove noise better but affects the edge and other fine details. Because of These drawbacks, we require improvement in standard median filters. Therefore lots of improved median filter has been developed like special median filter, Progressive switching median filter, rank ordered median filter, adaptive median filter etc.

3.1 Progressive Switching Median Filter

This filter is designed to remove highly corrupted images. First the filter read the image and finds the corrupted pixels [8]. The absolute difference of corrupted pixel from the median exceeds a predefined threshold value in an iterative manner. These iterations are performing up to certain number of times and only some particular neighbours are include for the median value. This above process will perform only when the flag of current iteration is different from the previous one, otherwise the pixel from previous image will be retained.

3.2 Rank Order Based Adaptive Median Filter

It is very efficient method of image de-noising [10]. In this method we first compare the value of current pixel with minimum and maximum value of the window and then compare the median value with minimum and maximum value of window. If the pixels and median is between these minimum and maximum values, then pixels will retain with its original value. If the median of window is between minimum and maximum value of the window but pixel is not between these values, then pixel will be replace by median value. If both median value and pixel values are out of minimum and maximum values limit then, size of the pixel window is increased & then calculate the median and replace that pixel with median value. If again the median is out of range of minimum and maximum values, then again increase the size of window up to some fixed maximum level. If the median is still out of the range of minimum and maximum value, then pixel left unchanged. The limitation of this approach is that noise free pixels and pixels corrupted by impulse noise, both are considered for determining the median value of window resulting in patches to occurs in images, especially when the noise ratio increase beyond 30 % or more and salt and pepper noise are unevenly distributed.

3.3 Adaptive Weighted Median Filter

This process contain three parts, the first one is noise detection over the image by taking average value of all pixels inside the window [11]. If central pixel's gray value = X(i,j)

Then all pixel inside the filtering window can be define as Y(i,j) = {X(i + k, j + r)},

$$\text{Where } \mathrm{k, r} = -1,\, 0,\, 1$$

And average of all pixels will be calculated as

$$\mathrm{Avg}\{\ \mathrm{Y(i,j)}\}\ = 1/9 \sum_{k=-1}^{1} \sum_{r=-1}^{1} 1\mathrm{X(i + k, j + r)} \tag{3}$$

If Min{Y(i,j)} = X(i,j) or Max{Y(i,j)} = X(i,j).
Then pixel is considered as noisy.

And we change the size of window according to the noisy pixels inside the window. After that noisy pixels will be filtered using the difference between central pixel & neighbour pixels.

4 Proposed Method

We have developed the simple algorithm in which we perform the noise detection & noise removal process simultaneously. We use the smallest window size which preserves the fine details of image. The window of size 3 × 33 × 3 chooses for noise detection and noise removal. The window contains total 9 elements which are as follows: Z1, Z2, Z3, Z4, Z5, Z6, Z7, Z8, Z9.

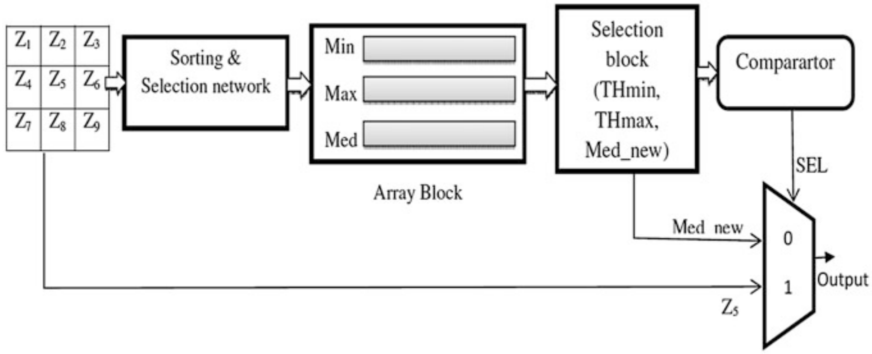


Fig. 1 Hardware representation of proposed method

Table 1 Filtering window of size 3 × 3

	Column 1	Column 2	Column 3
Row 1	Z_1	Z_2	Z_3
Row 1	Z_4	Z_5	Z_6
Row 3	Z_7	Z_8	Z_9

The hardware structure of complete process is shown in Fig. 1. Different blocks are shown for different steps. First block selects the maximum, minimum and median values of columns. Second block stores these values and third block selects minimum threshold, maximum threshold and final median value. Comparator use threshold values for noise detection and multiplexer uses final median value for noise removal.

We can divide the complete process into no. of steps as follows (Table 1):

Step-1 First we select all columns of filtering window one by one and then we find three values i.e. Maximum, Minimum and Median in each column. The mathematical expression can be shown as follow:

The minimum values of three columns are represented as

$$\text{Min }(\text{cln1}) = \min\{Z_1, Z_4, Z_7\}$$

$$\text{Min }(\text{cln2}) = \min\{Z_2, Z_5, Z_8\}$$

$$\text{Min }(\text{cln3}) = \min\{Z_3, Z_6, Z_9\}$$

The maximum value of these three column are represented as

$$\text{Max }(\text{cln1}) = \max\{Z_1, Z_4, Z_7\}$$

$$\text{Max }(\text{cln2}) = \max\{Z_2, Z_5, Z_8\}$$

$$\text{Max }(\text{cln3}) = \max\{Z_3, Z_6, Z_9\}$$

The median value of the these three column are represented as

$$\text{Med (cln1)} = \text{med}\ \{\ Z_1,\ Z_4,\ Z_7\}$$

$$\text{Med (cln2)} = \text{med}\ \{\ Z_2,\ Z_5,\ Z_8\}$$

$$\text{Med (cln3)} = \text{med}\ \{\ Z_3,\ Z_6,\ Z_9\}$$

Step-2 Now we have total nine values (three maximum, three minimum and three median). We will use these values to calculate threshold values (maximum threshold and minimum threshold) and median value. For these calculations, we make three different groups of these nine elements.

$$\text{Max\_group} = \{\text{Max (cln1)},\ \text{Max (cln2)},\ \text{Max (cln3)}\}$$

$$\text{Min\_group} = \{\text{Min (cln1)},\ \text{Min (cln2)},\ \text{Min (cln3)}\}$$

$$\text{Med\_group} = \{\text{Med (cln1)},\ \text{Med (cln2)},\ \text{Med (cln3)}\}$$

First we will calculate maximum threshold by choosing maximum value in min_group. Then we choose minimum threshold by choosing minimum value in max_group.

$$\text{Thmax} = \text{Max}\ \{\text{Min (cln1)},\ \text{Min (cln2)},\ \text{Min (cln3)}\}$$

$$\text{Thmin} = \text{Min}\ \{\text{Max (cln1)},\ \text{Max (cln2)},\ \text{Max (cln3)}\}$$

These two threshold values will be used for noise detection. Now we will calculate the median value for noise removal.

$$\text{Med\_new} = \text{Med}\ \{\text{Med (cln1)},\ \text{Med (cln2)},\ \text{Med (cln3)}\}$$

Step-3 Now we will perform noise detection and noise removal operation using these three values i.e. Thmax, Thmin, and Med_new. We compare the central pixel with threshold values. If the central pixel is in between the Thmin and Thmax, then the pixel will be considered as noise free, then pixel will remain unchanged and window will move or slide to the next pixel. Otherwise pixel will consider as noisy and it will be replaced by median value.

If

$$\text{Thmin} \leq \text{Z5} \leq \text{Thmax}$$

Then

$$X_5 \text{is unchanged.}$$

Else

$$X_5 = \text{Med\_new},$$

Table 2 Comparison of PSNR values of different filters for Lena image

De-noising Methods	Noise density				
	50 %	60 %	70 %	80 %	90 %
MF	14.99	12.19	9.80	8.10	6.50
CWM	12.99	10.80	8.90	7.60	6.29
PSMF	20.00	15.19	11.09	8.29	6.39
IMF	23.89	21.20	16.59	12.10	8.00
SDROM	14.40	11.70	9.40	7.80	6.40
ACWM	14.80	12.10	9.70	8.10	6.50
ACWMR	20.79	18.10	15.00	11.80	8.10
TSM	12.70	10.40	8.40	7.10	6.00
Proposed mETHOD	20.74	19.38	17.91	16.94	15.95

Table 3 Comparison of MSE values of different filters for Lena image

De-noising Methods	Noise density				
	50 %	60 %	70 %	80 %	90 %
MF	2057.7	3919.5	6808.9	10071.1	14557.2
CWM	3258.9	5408.5	8376.8	11300.0	15243.3
PSM	650.25	1963.9	5048.5	9619.0	14913.9
IMF	264.9	493.1	1422.6	4009.4	10305.7
SDROM	2360.9	4396.2	7465.8	10791.4	14896.3
ACWM	2153.1	4009.4	6967.5	10071.1	14557.2
ACWMR	540.8	1007.1	2056.2	4296.1	10071.1
TSM	3492.0	5930.3	9398.9	12678.8	16333.5
Proposed method	548.3	750.0	1052.1	1313.3	1650.3

Here we are parallelly calculating the threshold values and median value. So there is no need to perform noise detection and noise removal separately.

5 Simulation and Results

This method is mainly used for high density noise because most of the algorithms produce good results at low noise density but very poor results at high noise density. Using this method, we have performed image de-noising on Lena image of size 256 $\times$ 256 and simulate results on MATLAB 7.5. For different high noise density levels (50–90 %), the resultant PSNR is shown in Table 2, and we made a comparative analysis based on this peak signal to noise ratio (PSNR) of de-noised image. The results shown that, with the same noisy density the proposed filter gives much better PSNR then other median filters.

Fig. 2 Experimental Results of Proposed Method for Lena Image **a** original lena image **b** 50 % noisy **c** de-noised image **d** 60 % noisy **e** de-noised image **f** 70 % noisy **g** de-noised image **h** 80 % noisy **i** de-noised image **j** 90 % noisy **k** de-noised image

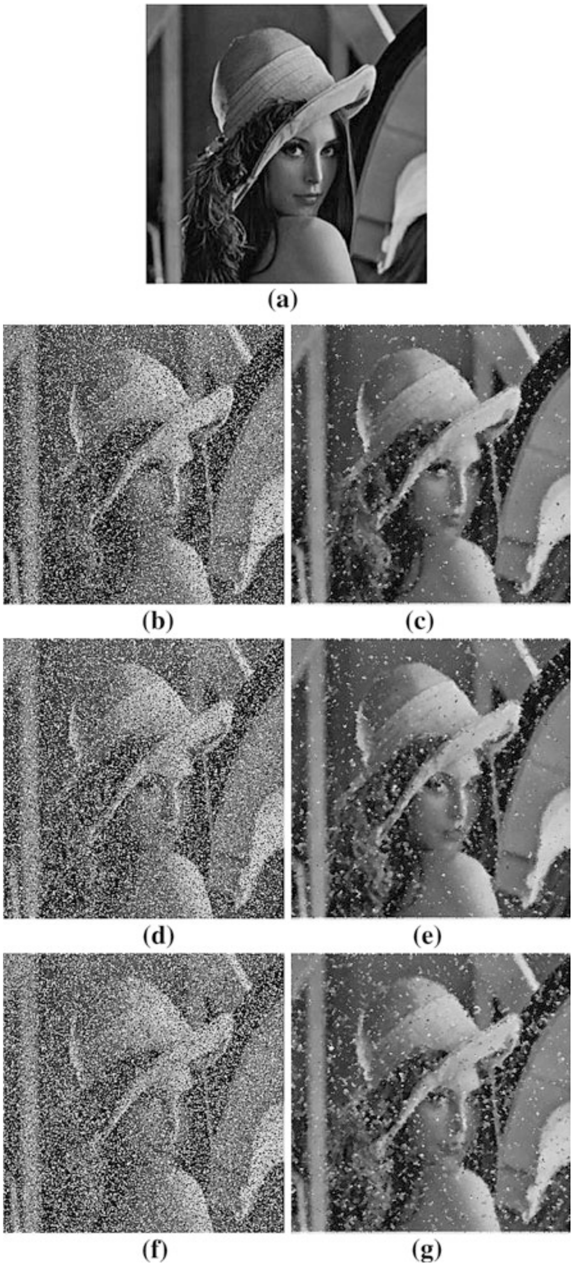
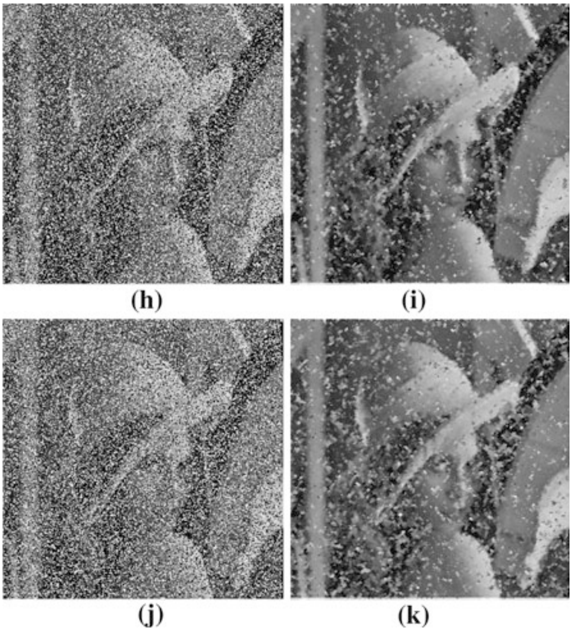


Fig. 2 continued



For the de-noised image, Z of size M × N, the PSNR [2] will be

$$PSNR = \ 10\log_{10}\frac{(255)^2}{\text{MSE}} \tag{4}$$

Where MSE (Mean square error), is

$$MSE = \frac{\sum_{i=1}^{m}\sum_{j=1}^{n}\{Z(i,j) - A(i,j)\}^2}{m \times n} \tag{5}$$

With respect to the noise-free original image A.

In the Table 2 shown below, we compared the PSNR and in the Table 3, MSE of different median filters i.e. Median filter (MF) [1], Centre weighted median filter (CWM) [4], Progressive switching median filter (PSMF) [8], Iterative median filter (IMF)[12], Signal dependent rank order median filter (SDROM) [13], Adaptive centre weighted median filter (ACWM) [10, 6], Recursive adaptive centre weighted median filter (RACWM), Tri-state median filter (TSM) [14].

The results in the Table 2 clearly show that the PSNR of proposed method is much better at high density of noise. As the density of noise increasing, the response of proposed filter is becomes better in comparison of other filters. Table 3 shows the comparative analysis of Mean Square Error (MSE) of different filters. The MSE of proposed filter is very lass as compare to other filters especially when the density of noise is very high. This method is tested on Lena image shown in Fig. 2. The Fig. 2b, d, f, h, j shows Lena image corrupted by 50, 60, 70, 80 and

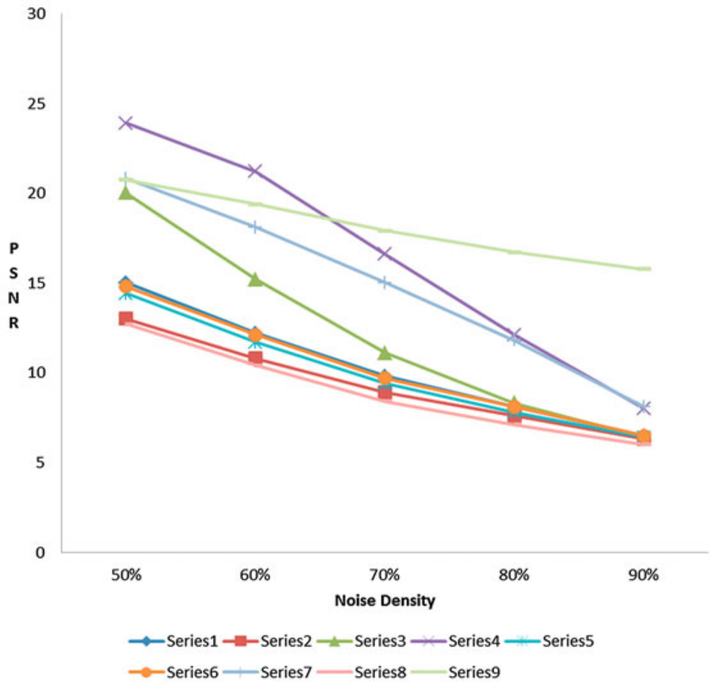


Fig. 3 Graphical representation of PSNR of different de-noising methods at different noise density

90 % respectively and Fig. 2c, e. g, i, k show images De-noised by proposed method. The Graphical representation of PSNR values of different filters shown in Fig. 3. It is very clear by the figure that PSNR of proposed method does not decrease very rapidly for high density noise like other filtering methods.

6 Conclusion

The proposed filter has proved that it is very efficient for random valued impulse noise because practically noise is not uniform over the channel. We have used the concept of maximum and minimum threshold to detect both positive and negative noise. It produces very good PSNR (Peak Signal to Noise Ratio) and very small MSE (Mean Square Error) for highly corrupted images, especially for more than 50 % noise density. This algorithm is simple and requires less number of calculations than other filters like CWM, TSM, SD-ROM, IMF etc. small size of filtering window gives advantage of preservation of fine details of image. Because of its less complexity of calculation, this filter will have great application in the field of image processing.

References

1. Gonzalez RC, Woods RE (2009) Digital image processing, 3rd edn. Pearson Prentice Hall, New Jersey
2. Nallaperumal K, Varghese J, Saudia S, Krishnaveni K, Mathew SP, Kumar P (2006) An efficient switching median filter for salt & pepper impulse noise reduction. In: 1st international conference on digital information management, 2006
3. Zhu Y, Huang C (2011) An improved median filtering algorithm combined with average filtering. IEEE third international conference on measuring technology and mechatronics automation, vol 6–7. Jan 2011, pp 420–423
4. Song T, Gabbouj M, Neuvo Y (1994) Center weighted median filters: some properties and applications in image processing. Signal Process 35(3):213–229
5. Chen T, Wu H (2001) Adaptive impulse detection using centre-weighted median filters. Signal Process Lett 8(1):13
6. Lin T-C, Yu P-T (2007) A new adaptive center weighted median filter for suppressing impulsive noise in images. Inf Sci 177:1073–1087
7. Nallaperumal K, Varghese J, Saudia S, et.al. (2006) Selective switching median filter for the removal of salt & pepper impulse noise. I: Proceedings of IEEE WOCN 2006, Bangalore, India
8. Wang Z, Zhang D (1999) Progressive switching median filter for the removal of impulse noise from highly corrupted images. IEEE Trans Circ Syst II: Analog Digital Signal Process 46(1):78–80
9. Dong F, Fan H, Da Y (2010) A novel image median filtering algorithm based on incomplete quick sort algorithm. Int J Digital Content Technol Appl 4(6):79–84
10. Xu X, Miller EL, Chen D, Sarhadi M (2004) Adaptive two-pass rank order filter to remove impulse noise in highly corrupted images. IEEE Trans Image Process 13(2):238–247
11. Chang-You W, Fu-ping Y, Hui G (2010) A new kind of adaptive weighted median filter algorithm, In: International conference on computer application and system modeling (ICCASM 2010)
12. Forouzan AR, Araabi BN (2003) Iterative median filtering for restoration of images with impulsive noise. In: Proceedings of the 10th IEEE international conference on electronics, circuits and systems, 2003
13. Abreu E, Mitra SK (1995) A signal-dependent rank ordered mean (SD-ROM) filter. A new approach for removal of impulses from highly corrupted images. In: Proceedings of IEEE ICASSP-95, Detroit, MI, 1995, pp 2371–2374
14. Chen T, Ma K–K, Chen L-H (1999) Tri-state median filter for image denoising. IEEE Trans Image Process 8(12):1834–1838
15. Behrooz G, Hadi SY, Faranak HY (2009) Modified adaptive center weighted median filter for suppressing impulsive noise in images. Int J Res Rev Appl Sci 1(3):218–228

Resting State Networks Analysis Using Simultaneous EEG-fMRI for Epilepsy Patient

Rajanikant Panda, Rose Dawn Bharath, Sandhya Mangalore, Neeraj Upadhyay, A. Thamodharan and Silpa Kanungo

Abstract The resting state EEG-fMRI has opened a new avenue in not only neuro cognitive studies but it has also found practical utility in clinical applications. We studied the Resting State Networks on Epilepsy Patient to understand the neuronal substrates involved in epilepsy. Five epilepsy patients were undertaken for simultaneous EEG-fMRI study. EEG microstates was computed and was considered as explanatory variables in the GLM design for the analysis of fMRI data in an event related design. z-stats and independent component was examined for simultaneous EEG-fMRI. We hypothesized that it's possible to analyze the affected brain areas for epileptiform discharges in epileptic patients at resting state. Microstates convolved functional image and its independent components using hybrid technique including both the neuronal and hemodynamic information was demonstrated on patients structural image. From this result we conclude that using EEG microstate and Independent Component Analysis (ICA) of resting fMRI we may examine the brain areas involved in resting state brain discharge. Also it will

R. Panda · R. D. Bharath (✉) · S. Mangalore · N. Upadhyay · A. Thamodharan · S. Kanungo
Cognitive Neuroscience Center, National Institute of Mental health and Neuroscience, Bangalore, Karnataka, India
e-mail: drrosedawn@yahoo.com

R. Panda
e-mail: bk.bme.rajanikant@gmail.com

S. Mangalore
e-mail: drsandym@gmail.com

N. Upadhyay
e-mail: neer86upadhyay@gmail.com

A. Thamodharan
e-mail: hamodharan@gmail.com

S. Kanungo
e-mail: silpabme2010@gmail.com

Mohan S. and S. S. Kumar (eds.), *Proceedings of the Fourth International Conference on Signal and Image Processing 2012 (ICSIP 2012)*, Lecture Notes in Electrical Engineering 222, DOI: 10.1007/978-81-322-1000-9_56,

be useful for the analysis of EEG-fMRI data in which electrical epileptic discharge are not apparent on scalp EEG at the time of data acquisition.

Keywords EEG · fMRI · GLM · Resting State Network · EEG-microstate · ICA

1 Introduction

Studying the model of networks involved in epilepsy help us understand the underlying mechanisms involved in triggering seizures [1]. Most of the patients with epilepsy may benefit if the exact location of the irritative zone can be established for further surgical interventions. Functional magnetic resonance imaging (fMRI) is non-invasive imaging technique and it allows the examination of brain structure and function with high spatial resolution but it lacks the temporal resolution at the level or speed at which cognitive neuronal processing occurs. Electroencephalography signal provides neuronal information with high temporal resolution and no doubt EEG is able to provide information on the subjects' vigilance and to some degree the state of mind but lacks the spatial information at the level of sub cortical neuronal substrates such as the basal ganglia. One possible solution is to combine the two measurements to map the changes in neural activity associated with epileptic discharges on to MRI images of brain anatomy [13].

Combined EEG-fMRI has indeed been demonstrated to be a valuable tool to delineate the irritative zone and is defined as the area with altered BOLD during epileptiform activity [12, 14, 15]. The analysis of simultaneous EEG-fMRI data is based on the identification of IED on EEG which are used for General Linear Model (GLM) of the fMRI analysis to localize IED generators but still the analysis is questionable if electrical epileptic discharge not occur or very few are detected in EEG at the time of fMRI scanning. Furthermore in many cases epileptic activity originating from sub cortical structures which cannot be recorded on the scalp EEG [2]. So to localize epileptic discharge when abnormality was not occurring in EEG signal or BOLD signal, alternative approaches are needed. Some studies reported where analysis was performed to identify the irritative zone even in the absence of interictal epileptiform discharges (IED) using simultaneous EEG-fMRI and post processing the EEG data into independent components and utilizing this as EEG regressor for the BOLD GLM model. The exact relation between neurovascular couplings has yet to been established. One studied repotted ICA was capable of highlighting areas of brain involved in epileptic activity in focal epilepsy patients [3, 7] even if IED was not occur at the time of EEG-fMRI data recording.

In this paper, we use simultaneous EEG-fMRI in patients with mesial temporal lobe epilepsy to lateralize the location of irritative zone using EEG microstates and by embedding this into the GLM of independent Component Analysis (ICA) of resting fmri [4]. We investigated the relationship between the presence of epilepsy-specific signature of the voltage maps on scalp electroencephalography with haemodynamic changes and whether this could help to localize the epileptic area. The spontaneous fluctuations of the EEG field maps (quasi-stable) was classified to some similar pattern of potential level. The neuroleptic episode was distinguished from resting brain electrical activity by classifying the slow and spike wave of IED as a separate quasi stable state from rest of normal neural microstates. This series of neuroleptic momentary potential distribution maps was clustered into successive time epochs ("microstates"). These EEG microstates were then used with BOLD resting state networks to examine the functional activity of brain for epileptic discharge. The z-state and independent components were examined to classify the different BOLD resting state signals and to identify the component which corresponded to the brain network involved in epileptic discharge.

2 Material

We included five adult right-handed epilepsy patients with a mean age of 25 years (SD $\pm$ 8 years) who were referred for clinical fMRI scanning to the NIMHANS Hospital, Bangalore, India. They have Mesial temporal lobe epilepsy. From video EEG report all the patients' epilepsy discharge was bilateral temporal lobe dominated to right temporal lobe.

3 Data Recording

Simultaneous resting EEG-fMRI was obtained. The data was recorded during eyes-closed, relaxed and awake condition. The subject was instructed as follows: "Please close your eyes and be relaxed and not to fall asleep". We used sound proof ear phones and maintained a delay of 300 ms in the MRI scanner so that the subject could habituate to the MRI scanner noise.

The exclusion criteria were if the subjects were uncooperative either due to claustrophobia or due to inability to follow the instructions given. Also if technical problems such as incorrect ECG recording was noted the subjects were discarded as this meant difficulty in cardioballistic artifact correction.

3.1 EEG Data Acquisitions

EEG data were recorded using a 32-channel MR compatible EEG system (Brain Products, Gilching, Germany). The EEG cap consisted of 31 scalp electrodes placed according to the international 10–20 system electrode placement and one additional electrode dedicated to the ECG. Data were recorded relative to an FCz reference and a ground electrode was located at Iz (10–5 electrode system [16]. Data were sampled at 5000 Hz and the impedance between electrode and scalp was kept below 5 kΩ. EEG was recorded using the Brain Recorder software (Version 1.2, BrainProducts).

3.2 fMRI Data Acquisition

Resting Functional MR-images were acquired using a 3T scanner (Skyra, Siemens, Erlangen, Germany). Echo-Planar Images (EPI) using BOLD contrast was obtained, 185 volumes were obtained applying the following EPI parameters: 34 slices, 6 mm slice thickness without any interslice gap, FOV 192 $\times$ 192 mm, matrix 64 $\times$ 64, repetition time 3000 ms, echo time 35 ms, refocusing pulse 90°, matrix- 256 $\times$ 256 $\times$ 114, voxel size-1 $\times$ 1 $\times$ 1 mm. We also acquired T1 MPRAGE data for better localize of patient brain activity.

4 Data Analysis

4.1 EEG Signal Analysis

4.1.1 Artifact Removal from EEG Signal

Raw EEG data were processed offline using BrainVision Analyzer version 2 (Brain Products, Gilching, Germany). Gradient artifact correction was performed using modified versions of the algorithms proposed by Allen et al. where a gradient artifact template is subtracted from the EEG using a baseline corrected sliding average of 20 MR-volumes. Data were then down-sampled to 250 Hz and low-pass filtered with an IIR filter with a cut-off frequency of 70 Hz. Following gradient artifact correction, the data were corrected for cardioballistic artifacts. The cardioballistic artifact is cardiac pulse artifact due to the static B0 field of the MR-tomography, this name is thought to be predominantly caused by cardiac-related body and electrode movement due to expansions and contraction of scalp arteries between the systolic and diastolic phase. Presumably to a lesser extent, there may also be fluctuations in the Hall-voltage (a potential which is created across a conductor with a current flow perpendicular to a surrounding magnetic field) due to the pulsating speed changes of the blood in the arteries. For this, an average artifact subtraction method was implemented in Brain Vision Analyzer2. This method involves subtracting the artifact on a second by second basis using heartbeat events (R peaks) detected in the previous 10 s. As such it requires accurate detection of R peaks which is aided by the employment of a moving average low pass filter and a finite impulse response high pass filter. In the present study, the R peaks were detected semi-automatically, with manual adjustment for peaks misidentified by the software. To average the artifact in the EEG channels, the R peaks are transferred from the ECG to the EEG over a selectable time delay. The average artifact was then subtracted from the EEG. Once gradient, cardio-ballistic had been removed, the data were then inspected visually for artifacts resulting from muscular sources or head movement artifact and any epoch containing those artifacts was rejected.

4.1.2 Contra Indication for MR and Cardiobalistic Artifacts

Always some artifacts may remain in the MR-EEG signal after correction of artifacts by cardiobalistic correction classifier, which conflict to examine the EEG and to extract the features. Especially the MR and Cardiobalisti artifacts which more or less looking as epilepsy spikes which may construct wrong EEG Microstate. Hence proper MR and Cardiobalisti artifacts artifact correction was needed for resting EEG-fMRI data processing. In addition for some patient/subject R peak was not coming properly in ECG graph, for this condition R peak suppose to be analyzes based on visually inspection.

4.1.3 Feature Extraction from Resting State EEG

Analyzing the resting EEG data is a big challenge. Previous studies have found independent component analysis and EEG-Microstates as the best way to describe resting brain electrical activity. The brain electric activity is a series of scalp maps of momentary potential distributions which change over time, and can be clustered into successive time epochs ("microstates") that are defined by quasi-stable landscapes of the brain electric field [17–20]. To compute the microstate the maxima of the Global Field Power (GFP) will be determined. Since topography remains stable around peaks of the GFP, they are the best representative of the momentary map topography in terms of signal to-noise ratio. The algorithm implemented for estimating the microstates, is based on a modified version of the classical k-means clustering method, in which cluster orientations are estimated (Pascual-Marqui et al. 1995). The optimal number of template maps was determined by means of a cross-validation criterion (Pascual-Marqui et al. 1995). We used this EEG microstate into the GLM design of resting fMRI data analysis at the single subject level. Extraction of these microstates was done using sLORETA software and further processed using a customized MATLAB program to obtain the individual microstates along with their onset time and duration. The microstates parameters thus obtained were then considered as explanatory variables (regressor) in the GLM design for the IC analysis of fMRI data.

4.2 fMRI Analysis

Our study aims to detect hemodynamic response during rest. The fMRI analysis was performed using FSL (FMRIB's Software Library, www.fmrib.ox.ac.uk/fsl). The first five functional image frames of each time series were discarded to allow for signal equilibration, giving a total of 180 frames used in analysis.

4.2.1 Pre Processing for Resting fMRI

The following pre-processing procedure was applied: employing different modules of the FSL-software package, we conducted motion correction using MCFLIRT [21], non-brain removal using BET [22], spatial smoothing using a Gaussian kernel of FWHM 5 mm, mean-based intensity normalization of all volumes by the same factor, and high temporal band pass filtering with sigma 90 s. These temporal filtering parameters were selected based on prior work demonstrating that spontaneous fluctuations upon which functional connectivity analyses are based exist in the range of 0.01–0.1 Hz [5]. After this we carried out ICA components analysis using MELODIC.

4.2.2 Visual Inspection of the Different IC for Artifact Identification

Since it was resting BOLD fluctuation and not task related activations, artifacts represent a challenging confounding factor, which often account for a large part of data variance, leading to a problematic data interpretation. A classic approach to deal with such confounds is to model them (e.g. motion-related noise) within the GLM [23]. Independent components were selected visually, by searching for anatomically relevant areas which could potentially depict functionally relevant resting-state networks. Components excluded from further analysis show clearly interpretable distinct artifact patterns such as motion related artifact or due to high spatial and temporal frequency noise or artifacts related to susceptibility artifacts by large vessels. We referred to the work of Beckmann et al. [24] and Beckmann and Smith [25] for identifying these "non motion-related" noise sources. The artifact components once identified were then filtered out using the MELODIC tool "reg_filt" (FMRIB's Software Library).

4.2.3 EEG Microstate Informed Resting-fMRI Analysis

The EEG microstates carried out as single trial "event related potentials (ERPs)" for the resting functional data analysis for examining IED events at the single subject level. In an event related design, EEG microstates was considered as an event and this was given as explanatory variables in the GLM model of resting fmri analysis. We modeled the input function using the onset time, and the duration of each EEG microstate and convolved them with a 3 column customized gamma hemodynamic response function. Then rerun the data for post-hoc regression analysis by inserting the GLM microstate information convoled with hemodynamic response function from feat analysis [6, 9, 11] (Fig. 1).

5 Resting State Network Examination

Different z-state and independent components of resting brain were obtained. The component showing the activations in the temporal lobe was identified based on the study by smith et al. The relation of the activity with respect to epileptic discharge was noted in all patients individually.

6 Result

Good quality EEG was obtained following gradient and pulse artifact subtraction, allowing identification of epileptiform discharges.

According to proposed method, inspection of all z-states and Independent components was carried out by expert observers revealed a component that qualitatively matched with region of interest. For evaluation of the result, we compare the result with standard modalities Video EEG.

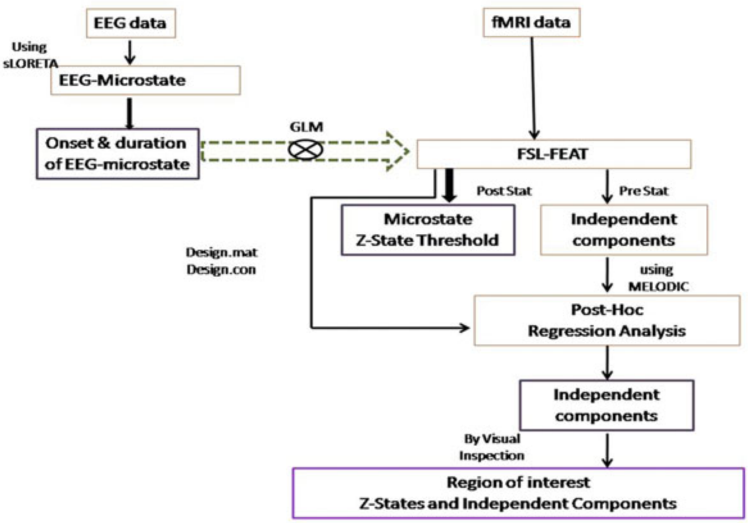


Fig. 1 Steps involved in the Z-Stats and Independent component processing

The following figure demonstrates the cortical signal change for epileptic discharge. The mean intensity projections of z-states and Independent components in single subject level.

Case 1.

In this case, according to video EEG, Patient had epileptic discharge on bilateral temporal region dominated to right medial temporal region (Fig. 2).

Case 2.

In this case, according to video EEG the epileptic discharge involving bilateral temporal lobe dominated to right anterior-medial temporal lobe, Right hippocampus region may be seizure onset zone (Fig. 3).

In summary, in all five temporal lobe (Mesial Temporal lobe Epilepsy) epilepsy patients the activated areas detected by proposed method is well concordant with areas defined in Video EEG during interictal recordings.

7 Discussion

Simultaneous EEG–fMRI is a non-invasive technique developed to improve the localization of IED generators with high spito temporal resolution in patients with epilepsy. If the interictal epileptic discharge was not occur at the time of data recording at that time it's a challenging to analyze the affected brain area for epileptic discharge. From past studies, its possible to analyze the affected brain area for epileptic discharge using ICA technique. But one of the prominent problems of ICA analysis is the potentially large number of components that require interpretation [26]. we assess the scalp quasic stability (EEG microstate) and ICA of fMRI to identify epileptic activity in patients with MTS epilepsy. We chose the independent components based upon the region of interest that is temporal lobe. We have cross validated the findings with long-term video EEG. In this study we examine five Epilepsy cases and analyzed them individually. In all five cases the activated areas detected by proposed method is well concordant with areas defined in Video EEG during interictal recordings.

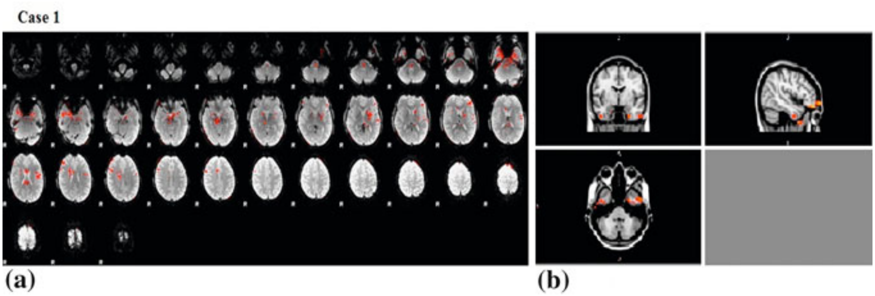


Fig. 2 EEG microstate and convolved them with a gamma hemodynamic response function showing bilateral temporal epileptic discharges. **a** z-state. **b** Independent component

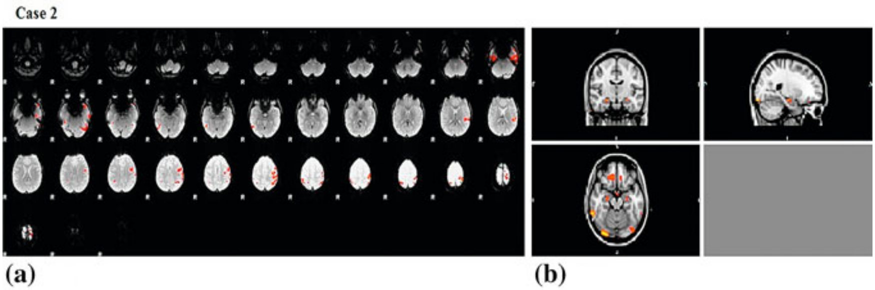


Fig. 3 EEG microstate and convolved them with a gamma hemodynamic response function showing bilateral temporal lobe epileptic discharges. **a** z-state **b** Independent component

This is a preliminary study highlighting the role of EEG-microstate and resting state fMRI analysis in patients with Epilepsy, and to analyze the data even when interictal epileptic discharge does not occur in EEG signal at the time of data recording. Future large studies are required to validate this technique. Comparison between the BOLD ICA components obtained with and without spikes using connectivity analysis can help us understand the dynamics of epilepsy and associated pathophysiology better.

8 Conclusion

Simultaneous EEG source imaging could improve the modeling of the BOLD signal in interictal studies of patients with epilepsy even if electrical epileptic discharge are not apparent on scalp EEG at the time of data acquisition, it's possible to localize and analyze the area of brain involved with epileptic activity in patients with MTS epilepsy. So EEG-fMRI is a useful tool to study the pathophysiological mechanisms of epilepsy and may assist in pre surgical evaluation of epilepsy. As the presented study results validated with the results of standard modalities Video EEG, still more ever validation and assessment of gains in sensitivity and specificity of the method needed, which should be addressed in larger groups of patients.

Acknowledgments We thank Dr. S. R. Chandra to examine the EEG signal and Dr. Shobini L. Rao, Dr. Arun Kumar Gupta for their inspiration and very helpful assessment on the simultaneous EEG-fMRI bring into play. Cognitive Neuroscience Center is supported by the Department of Science and Technology (DST), India.

References

1. Laufsa L, Hamandi K, Walkera MC, Scottb C, Smitha S, Duncana JS, Lemieux L (2006) EEG–fMRI mapping of asymmetrical delta activity in a patient with refractory epilepsy is concordant with the epileptogenic region determined by intracranial EEG. Magn Reson Imaging 24:367–371
2. Niedermeyer E, Lopes da Silva F (2004) Electroencephalography: basic principles, clinical applications, and related fields, 1st edn. Williams & Wilkins, Baltimore
3. LeVan P, Tyvaert L, Moeller F, Gotman J (2010) Independent component analysis reveals dynamic ictal BOLD responses in EEG-fMRI data from focal epilepsy patients. NeuroImage 49:366–378
4. Wang Z, Ives JR, Norton L, Hutchison RM, Mirsattari SM (2012) Spontaneous EEG-functional MRI in mesial temporal lobe epilepsy: implications for the neural correlates of consciousness. Epilepsy Res Treat 2012(385626):1–10
5. Biswal B, Yetkin FZ, Haughton VM, Hyde JS (1995) Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. Magn Reson Med 34(4):537–541
6. Britz J, Ville DVD, Michel CM (2010) BOLD correlates of EEG topography reveal rapid resting-state network dynamics. NeuroImage 52:1162–1170
7. Grouiller F, Thornton RC, Groening K, Spinelli L, Duncan JS, Schaller K, Siniatchkin M, Vulliemoz S, Lemieux L, Seeck M, Michel CM, Vulliemoz S (2011) With or without spikes: localization of focal epileptic activity by simultaneous electroencephalography and functional magnetic resonance imaging. Brain 156:1–20
8. Lei W, Eichele T, Calhoun VD (2010) Reactivity of hemodynamic responses and functional connectivity to different states of alpha synchrony: a concurrent EEG-fMRI study. NeuroImage 52:1252–1260
9. Musso F, Brinkmeyer J, Mobascher A, Warbrick T, Winterer G (2010) Spontaneous brain activity and EEG microstates. A novel EEG/fMRI analysis approach to explore resting-state networks. NeuroImage 52:1149–1161
10. Wang Z, Norton L, Hutchison RM, Ives JR, Mirsattari SM (2011) Spontaneous EEG-functional MRI in mesial temporal lobe epilepsy: implications for the neural correlates of consciousness. Epilepsy research and treatment, Article ID 385626
11. Yuan H, Zotev V, Phillips R, Drevets WC, Bodurka J (2012) Spatiotemporal dynamics of the brain at rest-Exploring EEG microstates as electrophysiological signatures of BOLD resting state networks. NeuroImage 60:2062–2072
12. Formaggio E, Storti SF, Bertoldo A, Manganotti P, Fiaschi A, Toffolo GM (2011) Integrating EEG and fMRI in epilepsy. NeuroImage 54:2719–2731
13. Ives JR, Warach S, Schmitt F, Edelman RR, Schomer DL (1993) Monitoring the patient's EEG during echo-planar MRI. Electroenceph Clin Neurophysiol 87:417–420
14. Krakow K, Wieshmann UC, Woermann FG (1999) Multimodal MR Imaging: Functional, diffusion tensor, and chemical shift imaging in a patient with localization related epilepsy 40:1459–1462
15. Salek-Haddadi A, Diehl B, Hamandi K, Merschhemke M, Liston A, Friston K, Duncan JS, Fish DR, Lemieux L (2006) Hemodynamic correlates of epileptiform discharges: An EEG-fMRI study of 63 patients with focal epilepsy. Brain Res. 88:148–166
16. Oostenveld R, Praamstra P (2001) The five percent electrode system for high-resolution EEG and ERP measurement Clin Neurophysiol 112:713–719
17. Brandeis D, Lehmann D (1989) Segments of ERP map series reveal landscape changes with visual attention and subjective contours. Electroencephalogr Clin Neurophysiol 73:507–519
18. Fingelkurts AA (2004) Making complexity simpler: multivariability and metastability in the brain Int J Neurosci 114:843–862
19. Michel CM, Murray MM, Lantz G, Gonzalez S, Spinelli L, Grave de Peralta R (2004) EEG source imaging, Clin Neurophysiol 115:2195–2222

20. Pascual-Marqui RD, Michel CM, Lehmann D (1955) Segmentation of brain electrical activity into microstates: model estimation and validation, IEEE Trans. Biomed. Eng 42:658–665
21. Jenkinson M, Bannister P, Brady J, Smith S (2002) Improved optimization for the robust and accurate linear registration and motion correction of brain images, NeuroImage 17:825–841
22. Smith S (2002) Fast robust automated brain extraction, Hum Brain Mapp 17:143–155
23. Friston KJ, Williams S, Howard R, Frackowiak RS, Turner R (1996) Movement related effects in fMRI time-series, Magn Reson Med 35:346–355
24. Beckmann CF, DeLuca M, Devlin JT, Smith SM (2005) Investigations into resting-state connectivity using independent component analysis, Philos Trans R Soc Lond B Biol Sci 360:1001–1013
25. Beckmann CF, Smith SM (2004) Probabilistic independent component analysis for functional magnetic resonance imaging, IEEE Trans Med Imag 23:137–152
26. Beckmann CF, Smith SM (2005) Tensorial extensions of independent component analysis for multisubject FMRI analysis, NeuroImage 25:294–311

Mobile Cloud Media Computing Applications: A Survey

S. Durga and Mohan S

Abstract The rapid developments of smart mobile devices, wireless networks and cloud computing have extended mobile phones with much more functionalities rather than only being used as voice communication tools. Smart mobile devices like camera phones are true "multimedia" devices capable of managing acquisition, processing, transmission, and presentation of multiple modal data such as image, video, audio and text information. They are also ubiquitous social networking devices. Recent rapid development of mobile hardware, wireless network, and cloud computing creates an urgent demand on intelligent and scalable multimedia computing technologies and systems for mobile devices. Multiple sensors, high-resolution screen, more powerful CPU, 3G Internet access, and cloud computing are all necessary enabling technologies to this end. In particular, cloud computing provides powerful supports on computation, storage and networking, which are essential for realizing scalable mobile multimedia applications. The motivation of this study is to timely address the challenges and opportunities in applying advanced multimedia technology, cloud computing, user interface, and graphics techniques to mobile systems. Existing works are reviewed, and an overview of recent advances in mobile cloud media computing is provided.

Keywords Cloud computing · Mobile computing · Mobile applications · Mobile cloud computing · Multimedia

S. Durga (✉)
Department of Information Technology, Karunya University, Coimbatore, India
e-mail: sdurga@karunya.edu

M. S.
Department of CSE, Dr. NGP Institute of Technology, Coimbatore, India
e-mail: s.mohan77@gmail.com

Mohan S. and S. S. Kumar (eds.), *Proceedings of the Fourth International Conference on Signal and Image Processing 2012 (ICSIP 2012)*, Lecture Notes in Electrical Engineering 222,
DOI: 10.1007/978-81-322-1000-9_57,

1 Introduction

Cloud computing is the delivery of computing as a service rather than a product, where by shared resources, software, and information are provided to computers and other devices as a utility over a network. This allows service providers and users to adjust their computing capacity depending on how much is needed at a given time or for a given task. Cloud computing and Mobile Computing domains have advanced rapidly and are the promising technologies for near future.

Cloud Computing [1] is a style of computing in which, typically, resources scalable on demand are provided "as a service(aaS) over the internet to users who need not have knowledge of, expertise in, or control over the cloud infrastructure that supports them. The provisional of the cloud services can be at the Infrastructural level (IaaS), Platform level (PaaS) or at the software level (SaaS). People are mainly using cloud storage services to accumulate their files such as images, videos, documents etc. and cloud processing services to reach their elastic needs for resources.

Similarly, improvements in the mobile devices, on hardware (memory, power consumption, touch screen, better ergonomic design, etc.), in software (numerous and sophisticated applications due to the release of iPhone [2] and Android [3] platforms) and in transmission (higher transmission rate with 3G and 4G technologies), have contributed in higher mobile penetration and better services provided to the customers. Future mobile devices may turn out to be mobile supercomputers as the integration of GPS, video camera, etc., with a decent battery lifetime [4]. With the rising perception of powerful mobile devices, mobile applications are emerging as a significant new application form in the next generation internet, among which media streaming may occupy the bandwidth of mobile devices. And the current wireless internet access technologies can only provide the choice between high average bitrate (e.g., WiFi or WiMAX) and good coverage area (e.g., cellular networks), but cannot guarantee both at the same time [5]. At the same time cloud computing is a promising technology which needs good applications for its large scale implementation.

Mobile technologies are also drawing their attention to the clouds due to the increasing demand of the applications, for processing power, storage space and energy saving. This has led to the Mobile Cloud Computing (MCC) domain. Applications that benefit from such a Mobile Cloud are from different domains like social networks, location based services, context-aware systems etc. Mobile cloud computing is defined as an extension of cloud computing in which mobile devices are the foundation hardware. It refers to an infrastructure where both the data storage and the data processing happen outside of the mobile device. Mobile computing means using portable devices to run stand-alone applications or accessing remote applications via wireless networks. With number of smartphones, tablets and other mobile devices are used every day, more and more users are relying on the cloud as the main driver for satisfying their need. In mobile cloud computing was first referred to as an infrastructure where data storage and processing could happen outside the mobile device. Mobile cloud applications move computing power and data storage away from mobile phones and into the cloud. This brings mobile applications and computing not just to smartphone users but to a broad range of mobile subscribers.

MCC can be divided into two. First one is, data storage and processing are take place outside the mobile devices, and it is the most convenient way of accessing the cloud services. One of the importance's of this approach is centralized security and maintains system. The second approach is data processing and storage is held within the mobile system.

MCC allows users to accept the services of cloud in dynamic environment. Mobile applications can be of two types: online applications and offline applications. In offline applications mobile devices are thick client that process the data locally on mobile devices with the data downloaded from the backend systems. In this type of applications there exists a periodic synchronization between the client and backend system. Online applications are based on the assumptions that a connection between mobile devices and backend system is always available. In the following sections, various mobile cloud multimedia applications are explained.

2 Mobile Cloud Multimedia Applications

Recent advances in smart phone technologies have fuelled a new wave of user demands for rich mobile experience. Today's mobile users not only expect broadband connection wherever they go and interaction with each other via social network on the road, but also seek ubiquitous access to a wealth of media-based contents and services. Since mobile devices are inherently resource-limited, cloud computing is emerging as a promising technology to provide additional resources for many media-rich mobile applications. However, the synthesis between mobile media and cloud computing should be well orchestrated to address many technical challenges arising in this exciting space.

The fundamental tension between resource-hungry multimedia streams and power-limited mobile devices has yet to be resolved, and is complicated by novel ways of operating mobile devices as both media clients and content providers. Efforts for providing a universal rich-media experience across any screen is typically hindered by the heterogeneity amongst ever-evolving mobile devices, as manifested in their different physical form factors, middleware platforms, and interactive features. In the era of cloud computing, the limitation on energy capacity can be eased off in an efficient way by offloading heavy tasks to the cloud. Here explains about the number of applications that are using the efficiency and performance of cloud with mobile computing. It includes the location identification, facial recognition, video learning, and cloud media system.

2.1 Location Identification

Nowadays all are using camera phones to capture and share their day to day experiences due to its mobility feature. Camera phones are capable of capturing, processing and transmitting of multimedia and textual information and other location services. The paper melog [6] stands for mobile/multimedia experience blogging where people can share their travel experience by employing mobile and the cloud services which reduce the manual effort of textual inputting. In melog, the existing efforts on mobile multimedia blogging are reviewed which can be categorized into three paradigms: manual photo blogging, text-free automatic blogging and text-rich automatic blogging. Manual blogging provides a feature for inputting the content of travel blog, and can upload this blog into their web site. Here the user needs to capture photos and input text to create a travel blog and this can be later uploaded. The disadvantages of this approach are that this does not provide sharing of real time experiences but only near realtime. And it is also a difficult task for users to input the text on mobile devices manually to publish the images on to the web.

The second approach, text-free automatic blogging avoids manual inputting of text. This uses the idea of space –timer relationship to display the photos on the web. For example, Space–Time Travel Blogging [7] displays photos on the map and shows their taken location, which are obtained from a GPS sensor. GeoLife [8] analyses GPS data and also shows photos taken during a trip on a map by time order. It provides a social networking service based on location as well. In the third approach text-rich automatic blogging, relatively friendly interface is used to help the users to create a travel blog automatically. Travelog [9] falls in this paradigm where the users can publish a travel blog by choosing photo from a camera phone and the related textual information, such as, annotation, weather and related links can be automatically mined from the internet. In melog they define a place as a spot of interest where the people are interest each spot of interest. The system architecture contains two parts: an application to take photos. The idea used in this approach is they build some sample photos on mobile devices and a services provided by the client. People can capture photos and videos in the mobile and this can be uploaded into a cloud platform. Next the micro blogging can be done in three modes fully automatic, semi-automatic and manual.

In fully automatic the user location can be recorded automatically with the help of geo location services and is sent to the location analyzer in the cloud to get information about the users moving status. When the location analyzer finds that the user had taken some photos then it will automatically generate a micro blog and this is uploaded into user's micro blog. In the semi-automated service some additional functions is provided to the users such as they can be able to review and if needed it is possible to edit the content and also they can decide whether to publish it or not. In the manual mode the user can manually trigger a micro-blog generation or publishing. The main advantage of this system is the blog generation will not affect the users' normal mobile operation. They can continue with their own work the blog can be automatically generated and this avoids the traditional way of manual text inputting and all. The main disadvantage of this system observed is the need of huge repository in the cloud.

Video Based Mobile Location Search with Large Set of SIFT Points in cloud [10] focuses on capturing the images of building and provides the summary of locations and sends to cloud and use service from the cloud. This technology comes into the existence due to increase in the penetration of mobile devices, PDAs etc. currently GPS and cell identification system is used. GPS system using the 24 satellite and cell identification is limited with the coverage. The identification process is done by feature extraction and comparing applying the SIFT algorithm on the data. The SIFT algorithm consist of local point extraction, key point localization, orientation assignment, key point descriptor. Data base consist of large amount of data such as videos and videos divided into number of frames. Using SIFT algorithm extract the SIFT primitives. Each SIFT primitive can be represented as a 135D vector, including video number, frame number, building number, 2D location, scale, orientation and 128D SIFT descriptor. After applying SIFT algorithm to the whole video database, a large-volume SIFT repository is obtained.

There are two part in the video repository the ground truth table for obtain the final result other part is purely128D descriptor. After obtains the SIFT primitives it pruning the primitives. Indexing can be done by using K-d tree and Best-Bin-First search are for better solutions. SIFT-base building matching for mobile location search could have a strong impact on many areas of m-commerce, such as, navigation aids for the visually impaired, tourist information/guide systems, etc. Future work goes in the direction of exploiting efficient indexing technique for high dimensionality data and geometric relations between SIFT descriptors to provide robust clustering to get a better performance. The other applications include mobile gaming. In mobile gaming, cloud help to searching services, and it provides the efficient way of searching the data and it uses the offloading energy. Mobile learning, mobile healthcare, M-commerce is the other applications.

Difficulties associated with these applications are secure storage of confidential data, processing delay, attack on the mobile devices, and energy efficiency of mobile devices. In the location identification methods, the captured image consist of dynamically moving object so it is very difficult to identify the actual location of the system. Also the cryptography method is needed to identify the user is valid person or not.

2.2 Human Tracking

An android application, CroudSTag [11] helps in forming a social group of common interest from mobile devices. The users of this application capture the images using mobile devices and store them into multiple clouds. Then the users can collect the images from the cloud and used to detect and identify the process users are log on to the facebook.com and collect their necessary information by selecting the cloud. Once the cloud is selected the mobile cloud computing started the face recognition process and mobile is notify about the result. Then the user sends an invitation to the other person to join a social group in Facebook (Fig. 1).

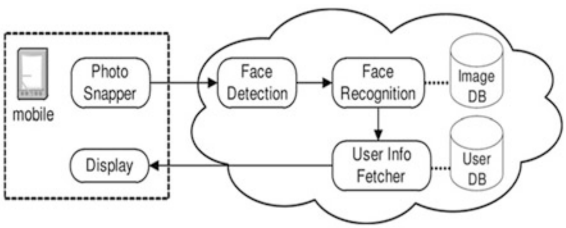
To construct the croudSTag application several services from cloud are consider such as facial recognition service from face.com. It offers three types of services: First photo finder, it scans the public photos in the network and allows tagging the photos. The second type is photo Tagger it allow people to scan the faces for grouping, third one is Celebrity finder it allow scans the pictures in twitter looking for celebrities. μCloud [12], a framework for rich mobile applications based on software composition is proposed for social face recognition application. It enables a user to automatically retrieve the profile of a face taken with an android device.

2.3 Mobile Learning Systems

Mobile learning is designed based on electronic learning and mobility. However, traditional mobile learning applications have limitations in terms of high cost of devices and network, low network transition rate, and limited educational resources [13]. Cloud based mobile learning applications are introduced to solve these limitations. So the most organizations are now using the advantages of cloud to access the data. Mobile learning system [14] deals with the interactive learning of the system with using the cloud technologies and mobile devices. In which the mobile user send request to the cloud for the video learning. The cloud provider buffer the data to the mobile system and user can also post the comment about that cloud computing provides the more secure, reliable data storage center, so large number of educational resources can be stored in the cloud server, making up for deficiencies of mobile devices with low storage capacity. In mobile learning the cloud plays an important role for the storage and security of data and cloud helpful for managing the load in peak hours.

In addition, [15] presents the benefits of combining mobile learning and cloud computing to enhance the communication quality between students and teachers. In this case, smartphone software based on the open source JavaME UI framework for clients is used. Through a web site built on Google Apps Engine, students

Fig. 1 Face recognition application introduced in [12]



communicate with their teachers at any time. Also, the teachers can obtain the information about student's knowledge level of the course and can answer students' questions in a timely manner. Another example of MCMC applications in learning is a contextual m-learning system based on IMERA platform [16] shows that a cloud-based m-learning system helps learners access learning resources remotely.

In [17], an education tool is developed based on cloud computing to create a course about image/video processing. Through mobile phones, learners can understand and compare different algorithms used in mobile applications (e.g., deblurring, de-noising, face detection, and image enhancement). The purpose of the deployment of these applications is to help the students enhance their understanding about the appropriate design of mobile cloud computing in supporting field experiences.

2.4 Cloud Streaming Applications

The pervasiveness of mobile devices and the Internet is changing the way how media information about our daily activities could be collected, accessed, processed and consumed. The accesses availability of rich media continues to accelerate in amount, variety, complexity and scale. This has been exemplified by Flickr, Facebook, YouTube and other online media sharing portal. However, lack of techniques for intelligent data processing and integration has become the major problem for effective management of data. Cloud Streaming promises to let resource constrained mobile devices handle compute intensive applications like interactive video streaming and gaming [18].

The paper E-Recall [19] explains the novel based cloud platform for the mobile rich data management. There are three main important concept (1) scalable media data processing, (2) flexible media content sharing and publishing and (3) personalized media content integration and search under mobile. The multimedia enrich data can be stored and processed in the cloud efficiently. For providing the intelligent processing and retrieval of data it provide the three concept, they are Multimodal based query modeling scheme, Cloud based database access method, User-centric rich media sharing/publishing module. In Query based modeling scheme it combines the different kind of query evidences. The processing have the two weakness first one is Incomprehensiveness second one is Instability and poor robustness. User-centric rich media sharing/publishing module template driven approach is developed via data driven approach. It aims to facilitate effective cloud based personalized rich media data sharing and publishing process under mobile environments.

In addition, A cloud assisted power efficient mobile P2P media streaming architecture [20] that addresses this problem of wireless access technologies, where clouds are responsible for storage and computing demanding tasks, and mobile devices collocating with each other share bandwidth and cooperatively share media content to distribute load.

3 Future Directions

As discussed in the previous section, MCMC has many advantages for mobile users and service providers. However, because of the integration of two different fields, i.e., cloud media computing and mobile networks, MCMC has to face many technical challenges. In addition, MCC would be deployed in a heterogeneous access scenario with different ratio access technologies such as GPRS, WLAN, and WiMAX [21]. Also MCMC requires wireless connectivity with the following features.

- MCC requires "always on" connectivity for a low data rate cloud control signaling channel.
- MCC requires an on demand wireless connectivity with scalable link bandwidth.
- MCC requires network selection and use that takes energy efficiency and costs into account.

Hence the critical challenge is to guarantee a wireless connectivity that meets the requirements of MCC with respect to scalability, availability, and energy and cost efficiency. Several research works contribute to the development of Mobile Cloud Media Computing (MCMC) by tackling issues as presented in the previous section. However, there are still some issues which need to be addressed. There is an increasing interest in sharing media files with family and friends. Leveraging cloud for efficient multimedia content management is further extending MCC into the research focus [22].

This section presents several open issues and possible research directions in the development of MCMC.

- Low Bandwidth—Although many researchers propose the optimal and efficient way of bandwidth allocation, the bandwidth limitation is still a big concern because the number of mobile and cloud users is dramatically increasing. Consider that 4G network is emerging as promising technologies that overcome the limitation and bring a revolution in improving bandwidth. However, the mobile users may face some problems such as congestion due to the limitation of wireless bandwidths, network disconnection, and the signal attenuation caused by mobile users' mobility. They cause delays when users want to communicate with the cloud, so QoS is reduced significantly. This problem is addressed in [20], and it proposes solution to share the limited bandwidth among mobile users who are located in the same area (e.g., a workplace, a station, and a stadium) and involved in the same content (e.g., a video file).
- Computational offloading issues—As explained in the previous section, offloading is one of the main features of MCMC to improve the battery lifetime for mobile devices and to increase the device performance. However, offloading is not always the efficient way to save energy [23]. In addition, determining which portions of the application's code need to be offloaded to improve the energy efficiency is a critical problem. The paper [24] suggests a program partitioning based on the estimation of the energy consumption before a program execution.

To find the optimal decision for partitioning applications before offloading is still a challenge.

- Quality of Service- in MCC, mobile users need to access to servers located in a cloud when requesting services and resources in the cloud. However, the mobile users may face some problems such as congestion due to the limitation of wireless bandwidths, network disconnection, and the signal attenuation caused by mobile users' mobility. They cause delays when users want to communicate with the cloud, so QoS is reduced significantly. The two new research directions are Clone Cloud [25] and Cloudlets [26] which are bringing the power of cloud computing to your smart phone through nearby high performance data centers.

4 Conclusion

In this day to day changing technology environment, demands of users also changes. Users demands quality multimedia service at anytime and anywhere with speed and accuracy. Mobile cloud media computing is one of the trends in future. It integrates the advantages of both the cloud computing and mobile computing for media processing. In this paper, few applications have been discussed which clearly show the applicability of the mobile cloud media computing to a wide range of mobile services. Then, the issues and related approaches for mobile cloud media computing have been discussed. Finally, the future research directions have been outlined.

References

1. Armburst M et al (2009) Above the clouds, a Berkeley view of cloud computing: technical representation. University of California
2. Apple, Inc, IPhone. http://www.apple.com/iphone
3. Android. http://www.android.com
4. Austin T, Blaauw D, Mahlke S, Mudge T, Chakrabarthi C, Wolf W (2004) Mobile supercomputers, computer. Proceedings of the IEEE, pp 81–83
5. Stiemerling M, Kiesel S (2009) A system for peer-to-peer video streaming in resource constrained mobile environments. U-NET′09
6. Li H, Hua X-S (2010) Melog—mobile experience sharing through automatic multimedia blogging. In: Proceedings of the ACM international workshop on Mobile cloud media computing, pp 19–24
7. Bamford W, Coulton P, Edwards R (2007) Space-time travel blogging using a mobile phone. In: Proceedings of the ACM international conference on advances in computer entertainment technology, p 8
8. Zheng Y, Chen Y, Xie X, Ma W (2009) GeoLife2. 0: a location-based social networking service. In: Proceedings of the international conference on mobile data management
9. Cemerlang P (2006) Towards automatic mobile blogging. In: Proceedings of the IEEE international conference on multimedia and Expo, pp 2033–2036

10. Ye Z, Chen X, Li Z (2010) Video based mobile location search with large set of SIFT points in cloud. In: Proceedings of the ACM international workshop on mobile cloud media computing, pp 25–30
11. Sriram SN, Paniagua C (2011) Huber CroudSTag: social group formation with facial recognition and mobile cloud services. In: Proceedings of the international conference on mobile web information system (MobiWIS), pp 633–640
12. March V et al (2011) μCloud: towards a new paradigm of rich mobile applications. In: Proceedings of the international conference on mobile web information systems, pp 618–624
13. Chen X, Zhai Y (2010) System design of cloud computing based on mobile learning. In: Proceedings of the international conference on knowledge acquisition and modeling. pp 293–242
14. Saranya SM, Vijayalakshmi M (2011) Interactive mobile live video learning system in cloud environment. In: Proceedings of the IEEE international conference on recent trends in information technology (ICRTIT), pp 673–677
15. Zhao W, Sun Y, Dai L (2010) Improving computer basis teaching through mobile communication and cloud computing technology. In: Proceedings of the IEEE international conference on advanced computer theory and engineering, pp 452–454
16. Yin C, David B, Chalon R (2009) Use your mobile computing devices to learn. In: Proceedings of the IEEE international conference on computer science and information technology, p 440
17. Ferzil R, Khalife I (2011) Mobile cloud computing educational tool for image/video processing algorithms. In: Proceedings of the IEEE signal processing education workshop, p 529
18. Lawton G (2012) Cloud streaming brings video to mobile devices. (IEEE) computer, pp 14–18
19. Shen J, Yan S, Hua XS (2010) The E-recall environment for cloud based mobile rich media data management. In: Proceedings of the ACM international workshop on mobile cloud media computing, pp 31–34
20. Jin X, Kwok Y-K (2010) Cloud assisted P2P media streaming for bandwidth constrained mobile subscribers. In: Proceedings of the IEEE international conference on parallel and distributed systems, pp 800–806
21. Guan L et al (2011) A survey of research on mobile cloud computing. In: Proceedings of the IEEE international conference on computer and information science, pp 387–392
22. Diaz D et al (2011) Media cloud: an open cloud computing middleware for content management. In: IEEE Trans Consum Electr 57:970–978
23. Rudenko A et al (1998) Saving portable computer battery power through remote process execution. Journal of ACM SIGMOBILE
24. Kumar K, Lu Y (2010) Cloud computing for mobile users: can offloading computation save energy. IEEE computer society
25. Chun G et al (2011) CloneCloud: elastic execution between mobile device and cloud. In: Proceedings of the international conference on computer systems, pp 301–314
26. Satyanarayanan M et al (2009) The case of VM based coudlets in mobile computing. IEEE pervasive computing pp 14–23

Author Index

Mohan S. and S. S. Kumar (eds.), *Proceedings of the Fourth International Conference on Signal and Image Processing 2012 (ICSIP 2012)*, Lecture Notes in Electrical Engineering 222,
DOI: 10.1007/978-81-322-1000-9, © Springer India 2013