Advances in Intelligent Systems and Computing 678

Sabu M. Thampi
Sri Krishnan
Juan Manuel Corchado Rodriguez
Swagatam Das
Michal Wozniak
Dhiya Al-Jumeily *Editors*

Advances in Signal Processing and Intelligent Recognition Systems

Proceedings of Third International Symposium on Signal Processing and Intelligent Recognition Systems (SIRS-2017), September 13–16, 2017, Manipal, India

Springer

Advances in Intelligent Systems and Computing

Volume 678

Series editor

Janusz Kacprzyk, Polish Academy of Sciences, Warsaw, Poland
e-mail: kacprzyk@ibspan.waw.pl

About this Series

The series "Advances in Intelligent Systems and Computing" contains publications on theory, applications, and design methods of Intelligent Systems and Intelligent Computing. Virtually all disciplines such as engineering, natural sciences, computer and information science, ICT, economics, business, e-commerce, environment, healthcare, life science are covered. The list of topics spans all the areas of modern intelligent systems and computing.

The publications within "Advances in Intelligent Systems and Computing" are primarily textbooks and proceedings of important conferences, symposia and congresses. They cover significant recent developments in the field, both of a foundational and applicable character. An important characteristic feature of the series is the short publication time and world-wide distribution. This permits a rapid and broad dissemination of research results.

More information about this series at http://www.springer.com/series/11156

Sabu M. Thampi · Sri Krishnan
Juan Manuel Corchado Rodriguez
Swagatam Das · Michal Wozniak
Dhiya Al-Jumeily
Editors

Advances in Signal Processing and Intelligent Recognition Systems

Proceedings of Third International Symposium on Signal Processing and Intelligent Recognition Systems (SIRS-2017), September 13–16, 2017, Manipal, India

Springer

Editors

Sabu M. Thampi
School of CS/IT
Indian Institute of Information Technology and Management
Trivandrum, Kerala
India

Sri Krishnan
Department of Electrical and Computer Engineering
Ryerson University
Toronto, ON
Canada

Juan Manuel Corchado Rodriguez
Department of Computer Science
University of Salamanca
Salamanca, Salamanca
Spain

Swagatam Das
Electronics and Communication Sciences Unit
Indian Statistical Institute
Kolkata, West Bengal
India

Michal Wozniak
Department of Systems and Computer Networks
Wroclaw University of Science and Technology
Wroclaw
Poland

Dhiya Al-Jumeily
Faculty of Engineering and Technology
Liverpool John Moores University
Liverpool
UK

ISSN 2194-5357 ISSN 2194-5365 (electronic)
Advances in Intelligent Systems and Computing
ISBN 978-3-319-67933-4 ISBN 978-3-319-67934-1 (eBook)
DOI 10.1007/978-3-319-67934-1

Library of Congress Control Number: 2017952864

Printed on acid-free paper

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

Preface

This edited volume contains a selection of refereed and revised papers originally presented at the third International Symposium on Signal Processing and Intelligent Recognition Systems (SIRS'17). The symposium was held in Manipal Institute of Technology, Manipal University, Manipal, India, during September 13–16, 2017. SIRS'17 provided a forum for the sharing, exchange, presentation, and discussion of original research results in both methodological issues and different application areas of signal processing, computer vision, and pattern recognition.

We would like to thank all authors for their contributions to the program and for their contributions to these proceedings. The technical program of SIRS'17 comprises of 41 papers (24 regular papers and 17 short papers). These papers were selected by the program committee with additional help from external expert reviewers from 111 submissions. Each of them was reviewed by two or more referees. The authors were asked to address each and every comment made by the referees for improving the quality of their papers.

We are also deeply grateful to the many people who volunteered their hard work to ensure this successful symposium. We would like to express our gratitude to the program committee and external reviewers, who worked very hard in reviewing papers and providing suggestions for their improvements. Many thanks go to all the chairs, and their involvement and support have added greatly to the quality of the symposium. We also wish to thank all the members of the Advisory Committee, whose work and commitment were invaluable. We would like to express our sincere gratitude to local organizing committees that have made this event a success. We would also like to express our thanks to the keynote speakers and tutorial presenters. The EDAS conference system proved very helpful during the submission, review, and editing phases.

We wish to express our sincere thanks to Thomas Ditzinger, Senior Editor, Engineering/Applied Sciences Springer-Verlag, and Janusz Kacprzyk, Series Editor, for their help and cooperation.

We hope these proceedings will serve as a valuable reference for researchers and practitioners in the related fields.

Sabu M. Thampi
Sri Krishnan
Juan Manuel Corchado Rodriguez
Swagatam Das
Michal Wozniaķ
Dhiya Al-Jumeily

Organization

Committee

Chief Patron

Ramdas M. Pai	Manipal University, India

Patrons

H.S. Ballal	Manipal University
H. Vinod Bhat	Manipal University
V. Surendra Shetty	Manipal University
Narayan Sabhahit	Manipal University
G.K. Prabhu	MIT, Manipal University
B.H.V. Pai	MIT, Manipal University

Honorary Chair

K.R. Rao	University of Texas at Arlington, USA

General Chair

Sri Krishnan	Ryerson University, Toronto, Canada

Program Chairs

Juan Manuel Corchado Rodriguez	University of Salamanca, Spain
Michal Wozniak	Wroclaw University, Warsaw, Poland
Dhiya Al-Jumeily	Liverpool John Moores University, UK
Swagatam Das	Indian Statistical Institute, Kolkata, India

Advisory Committee

Teodiano Freire Bastos Filho	Universidade Federal do Espírito Santo, Vitoria, Brazil
Janusz Kacprzyk	Polish Academy of Sciences, Poland
Sankar K. Pal	Indian Statistical Institute, Kolkata, India
Nallanathan Arumugam	King's College London, UK
P. Nagabhushan	University of Mysore, India
Soura Dasgupta	The University of Iowa, USA
Ronald R. Yager	Machine Intelligence Institute, Iona College, USA
Jamila Mustafina	Kazan Federal University, Russia
Selwyn Piramuthu	University of Florida, USA
El-Sayed El-Alfy	King Fahd University of Petroleum and Minerals, Saudi Arabia
Abir Hussain	Liverpool John Moores University, UK
Laszlo T. Koczy	Szechenyi Istvan University, Gyor, Hungary
Ngoc Thanh Nguyen	Wroclaw University of Technology, Wroclaw, Poland
David Zhang	The Hong Kong Polytechnic University, Hong Kong
Millie Pant	Indian Institute of Technology Roorkee, India
Naeem Radi	Al Khawarizmi University College, UAE
Ibrahim Al-Jumaili	Al-Anbar University, Iraq

Steering Committee Chair

Sabu M. Thampi	IIITM-Kerala, India

Organizing Chair

Hareesha K.S.	Manipal Institute of Technology (MIT) - Manipal University, India

Organizing Co-chairs

Ashalatha Nayak	Manipal Institute of Technology, Manipal University
Balachandra	Manipal Institute of Technology, Manipal University

Organizing Secretaries

Renuka A.	Manipal Institute of Technology, Manipal University
Preetham Kumar	Manipal Institute of Technology, Manipal University
Poornima PK	Manipal Institute of Technology, Manipal University

Organized by

MANIPAL INSTITUTE OF TECHNOLOGY
MANIPAL
A Constituent Institution of Manipal University

In association with

LIVERPOOL
JOHN MOORES
UNIVERSITY

Contents

Signal and Image Processing

Intelligent Recognition Techniques and Applications

Signal and Image Processing

Removal of BW and Respiration Noise in abdECG for fECG Extraction

Jeffy Joseph[1](✉), J. Rolant Gini[1], and K.I. Ramachandran[2]

[1] Department of Electronics and Communication Engineering,
Amrita School of Engineering Coimbatore, Amrita Vishwa Vidyapeetham,
Amrita University, Coimbatore, India
jeffyjoseph2007@gmail.com, j_rolantgini@cb.amrita.edu
[2] Center for Computational Engineering and Networking,
Amrita School of Engineering Coimbatore, Amrita Vishwa Vidyapeetham,
Amrita University, Coimbatore, India
ki_ram@cb.amrita.edu

Abstract. Electrocardiogram (ECG) signals are one of the most important diagnostic tools for any doctor, especially a cardiologist. It is important that the fetus present inside the abdomen undergoes a fetal ECG recording to assess the health of the fetus. Complications like disturbance because of movement of abdominal muscles are usually present during the recording and leads to the wrong diagnosis of the fetus ECG. In this paper, the signal in dispute had been altered in the proposed method so as to eliminate the wandering of the baseline, respiration noise and also expel the noise from other sources. The acquired abdominal ECG signal in a noninvasive manner had been considered for extracting the fetal ECG after eliminating the noise. The windowed zero mean method is used where the first step is segmentation. In segmentation, the abdominal ECG signal is divided into set of samples based on window size. Zero mean is applied across each of the windowed abdominal ECG signals to address the issue of baseline wandering and respiration noise. This is followed by the application of a bandpass filter to cancel the high-frequency noise component. This process results in an ECG signal that almost has no complications as present before. The fetal ECG signal that is procured using such a method is now easier to diagnose as compared to the acquired signal which contains noise. Thus, for a fetus, this can help in proper diagnosis. It is further noted that this method is very reliant on using and is lucid. It can be used to augment and alter signals where such complications arise in the field of medicine and clinical diagnosis.

Keywords: fetal ECG (fECG) · abdominal ECG (abdECG) · Baseline wander (BW) · Respiration noise · Windowed Zero Mean (WZM)

1 Introduction

The simplicity of ECG waves is what makes them easy to comprehend and understand when compared to other wave signals. From an ordinary heart check-up to the detection of an arrhythmia we need to use the ECG [1]. The disturbance in an ECG due to

S.M. Thampi et al. (eds.), *Advances in Signal Processing and Intelligent Recognition Systems*, Advances in Intelligent Systems and Computing 678,
DOI 10.1007/978-3-319-67934-1_1

various noise can cause the wrong diagnosis due to bereavement of information. Some problems that arise in ECG signal measurement are due to noise such as baseline wandering which is also usually accompanied by high-frequency noise components. Baseline wandering can arise due to respiration and sweat that affects the parameters of the electrode such as impedance. High-frequency noise components are caused by movement of the body during the procedure.

Welfare of the fetus is the prime concern in pregnancy. Though the fetal mortality rate [2] has decreased in developed countries, it is still a grave issue in developing and underdeveloped nations. Genetic diseases always pose a large looming threat on the fetus. Congenital malformations, membrane complications, malaria, and pre-eclampsia are some of the disease responsible for fetal deaths. In a study in England and Wales in the year 2007, it was found that congenital anomalies are the second highest cause of infant deaths [3]. Bradycardia [4] has led to the invention of the micropacemaker [5] which when used in a fetus helps forestall the need for exigent removal from its womb in case of any medical emergency.

There are many ways to monitor the fetus using techniques like ultrasound which is based on the usage of sensors. Non-stress test and contraction stress test are also other ways of externally monitoring the state of the fetus. Presumptuous methods like abdECG also help in relatively accurate measurement. Fetal ECG (fECG) is obtained using electrodes which are planted on the head of the fetus. Such a prying procedure is very cumbersome and invasive, which calls for the requisite for procedures that are non-invasive. Observation of the fECG can help in having an insight of ailments and anomalies that could be present in the fetus. One of the latest developments is the use of Raspberry Pi microcontroller to monitor fECG [6]. There is various noise usually present in the ECG that can arise from various sources such as interference of the power line, noise due to contact established between electrodes, electrosurgical noise, instrumentation noise and electromyography (EMG) noise that occurs due to muscle movement in the vicinity of the heart. Since the fetus is present near the abdomen, there is the possibility of much clamor that can occur due to the position and movement of the fetus. The abdomen is in an anterior part of the body which contains important organs like stomach and liver.

The abdominal muscles present abet in the process of breathing as adjunct support. During this process, there is disturbance because of the process of respiration as there is the movement of abdominal muscles present near the electrode. This neighboring signal gets acquired as a part of the reading and is more prominent when compared to the actual signal obtained from the abdomen. The noise created due to these such as high-frequency noise components due to muscle movement and the wandering of the baseline have to be addressed. Electromyography noise arises due to muscle movement that can affect the changes in fECG especially if it occurs adjacent to the heart of the fetus. These noises are inevitable and induce fallacious results. Wandering of the baseline can occur due to improper placement of electrodes and this results in the ECG wave wandering above and below the reference due to variation in amplitude. Some of these disturbances cause the signal to get lost amidst these undesirable perturbations. These problems need to be adhered to as the welfare of the fetus depends on it amidst pregnancy. This demands for elementary methods to deal with it. The paper has been fabricated in the following order: Sect. 2 provides an accord about the literature survey that gives requisite background information. This is followed by Sect. 3 which presents the method with which the problem had been approached. This is followed by Sect. 4 gives a discourse about the results achieved through the Windowed Zero Mean (WZM) method applied in this paper. The Sect. 5 gives the conclusion inferred from this method.

2 Literature Review

For a fetus, especially, the ECG signal is taken from the abdomen of the mother and as seen previously there is a lot of disturbance because of noise and this can lead to a misdiagnosis. The primary reason for this is that the ECG signal masquerades as noise in certain parts of the signal and only upon close observation it is realized that it is not noise but rather indispensable information that is being lost. The usage of such vital information can aid in augmenting the process of diagnosis. Extraction of respiration signal from ECG signal (EDR) is carried out using methods like single lead ECG [7] where EDR is estimated by cubic-spline interpolation; conductive textile electrodes which use the concept of instantaneous frequency estimation [8].

Removal of noise from Surface respiratory EMG signal is also effectuated using methods such as lean mean square widrow adaptive structure [9], butterworth filtering [10] and adaptive filtering [11]. The Heart Rate Variability method is proposed as a more accurate method when using the single lead ECG [7] whereas the conductive textile electrode [8] method indicates a closer alternation with the actual respiration signal. The lean mean square widrow adaptive structure [9] removes noise without the need for additional electrodes. In butterworth filtering method [10], the noise due to the electrode is nullified using a butterworth filter of corner frequency 20 Hz. Recursive least squares algorithm is used in the adaptive filtering method [11] to eliminate Power Line Interference and ECG noise. An EDR extracted signal accommodates plenty of noise and needs to undergo pre-processing in order to obtain a clean ECG. This incites the requisite for unsophisticated methods that need to be devised in order to deal with it. Some of the acclaimed methods that had been used to address this problem of the ECG until now are Empirical Mode Decomposition (EMD) [12] based on Hilbert-Huang transform [13], linear phase filtering [14], moving average filter [15] and discrete wavelet transform [16]. EMD [12] which is based on Huang transform [13] relies purely on data which might be a disadvantage in certain cases where the data alone will not suffice. Summation of Intrinsic mode functions is generated upon decomposition of the EMD [12] which is the motive of the EMD [12] method.

Linear phase filtering method [14] primarily focuses on reduction in the number of computations needed to achieve de-noising of signal and elimination of baseline wandering. This, however, can cause signal distortion as there will be delay variation created due to the propagation of the signal through devices which may lead to deprivation of crucial information. The moving average filter [15] is applied to the ECG and is smoothened using polynomial curve fitting. However, this requires pre-processing which might not be suitable for all cases. Discrete Wavelet Transform [16] is used to choose wavelets and the depth to attain the level of decomposition is also decided. Following this, the wavelet is shrunk using Empirical Bayes posterior median. The problem with a wavelet-based approach is that the accuracy of this method will not suffice and lead to bereavement of information. While all these methods are used to deal with ECG, none of them indicate a method to expel noise in a fECG. The Windowed Zero Mean (WZM) method proposed has ensured that there is no loss of information and uses a relatively lucid approach. Using MATLAB, this method ensures smooth de-noising and reinstatement of the original fECG signal without any compromise on the original signal.

3 Methodology

The current method is used to address the issues stated above such as Electromyography noise and baseline wandering. The procedure on how the method is ingrained is stated in the block diagram present below (Fig. 1). An abdECG signal which consists of N values from x(1) to x(N) is considered as input to the segmentation stage of the block diagram. In the first stage, the signal is segmented and segregated into different windows. The segmented signal comprises of N windows with each window of length n1. So, taking the first window, elements of the signal from x(1) to x(n1) are denoted as W1; the second window, elements of the signal from x(n1 + 1) to x(2 * n1 + 1) are denoted as W2; and so on until the Nth window where elements from x(k * n1 + 1) to x(N) are present. Then, zero mean of each segmented window on using the formula given below is taken to deprive the signal of the undesired deviation of baseline.

$$X(k) = W(k) - \text{mean}(W_k(:)) \tag{1}$$

Here, X(k) [where k is an arbitrary number] is the zero mean value obtained for a window W_k and W(k) is an element in the window. The same is to be done for all the windows before concatenation. The signals after application of zero mean are identified and grouped accordingly from X(1) to X(N). Concatenation of all these segmented windows gives the resultant signal. In the post-processing stage, the signal with elements from X(1) to X(N) is passed through a bandpass filter with lower cut-off

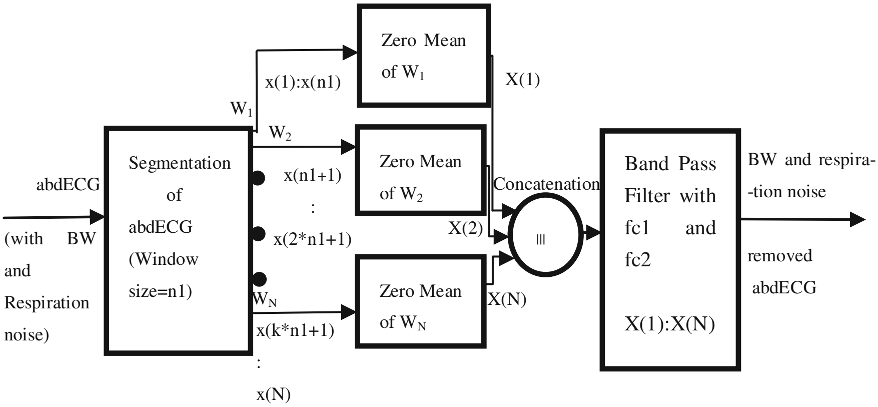


Fig. 1. Block diagram of Windowed Zero Mean Method

frequency 0.5 Hz and higher cut-off frequency 20 Hz to remove the high-frequency noise components. The output signal, X2, now obtained, consists of elements that are free from baseline wandering and high-frequency noise components.

3.1 Pseudo Code

```
Initialize parameters: window size, duration, sampling
frequency
Input: abdECG signal (x)
Segment x into k windows of window size n1
Do 1:k
Compute zero mean of each windowed signal
Enddo;
Concatenate: X(1): X(N)
Bandpass filter (fc1, fc2) of X
Output signal: X2
```

The database used for testing abdominal and fetal ECG's is obtained from Physiobank ATM [17, 18]. In the sample signal used, it is observed that the problem of baseline wandering and high-frequency noise constituents are significant and alter the details of the ECG, so it is proven that it is a favored signal to be tried and tested. The analysis of the first signal from the database is depicted. The first lead of the signal is seen in Fig. 2. It has a sampling frequency of 1000 Hz and the amplitude is measured in microvolts.

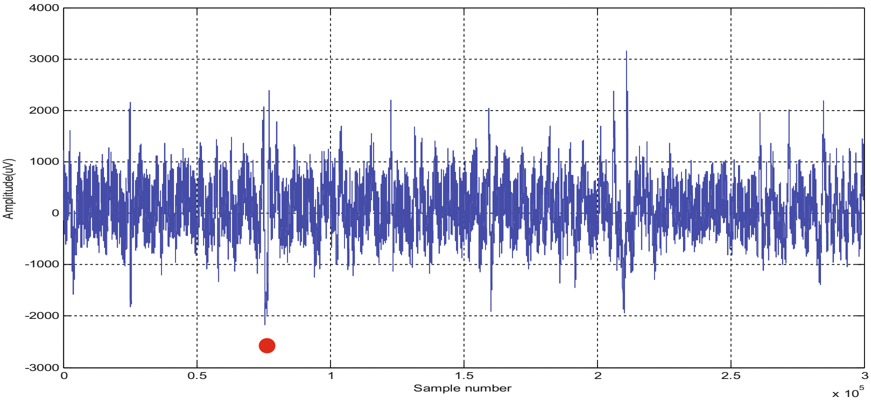


Fig. 2. abdECG with BW and respiration noise

It is indicated that there are baseline wandering and high-frequency noise components in the signal. Figure 3 depicts a closer view of the signal at that point.

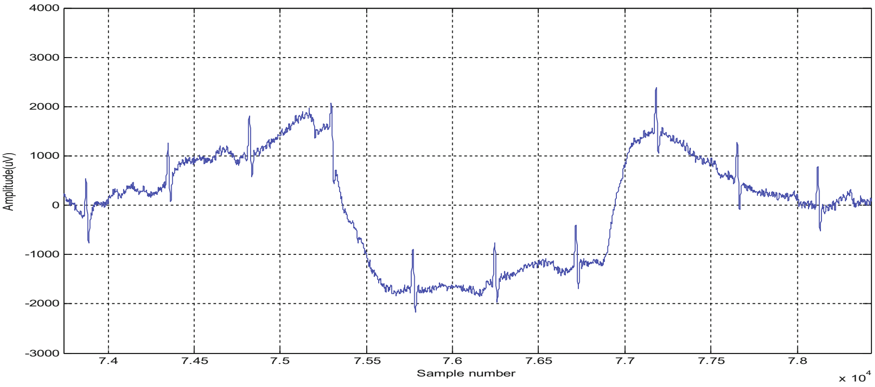


Fig. 3. Narrower view of abdECG indicating BW and respiration noise

In the first stage, the window size that is to be used is contemplated. In an adult, it had been observed that the average time taken to inhale and exhale followed by an involuntary pause is from 4 to 6 s and the number of breaths taken had been around 12–20 per minute [19]. The intake of oxygen during pregnancy that usually increases does not affect the respiration rate. Therefore it is seen that taking a sampling frequency of 1000 Hz construes the use of a large window size such as 1000. This, however, leads to erroneous results as shown in Fig. 4. It is observed that though the signal has undergone zero means, the acclimatized amplitude is still elevated and does not exhibit removal of baseline wander. Similarly, when a very small window size like 10 is used, the inherent original signal is lost and time for execution also is higher.

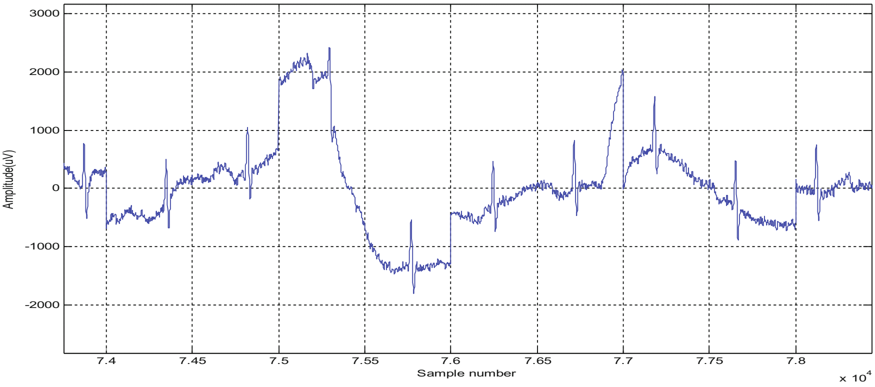


Fig. 4. Windowed Zero Mean (WZM) Signal with window size = 1000

This advocates the use of a smaller window size like 36 which yields finer results as the baseline wandering had been removed as denoted in Fig. 5.

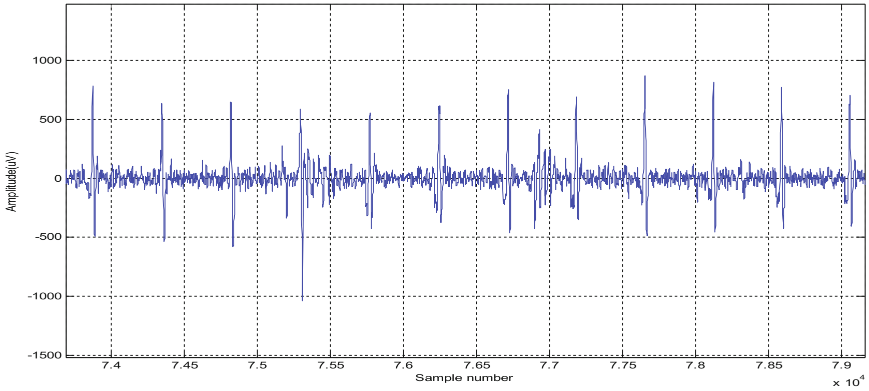


Fig. 5. Windowed Zero Mean (WZM) Signal with window size = 36

The signal is then segmented into different sets of samples based on the window size chosen. The next stage comprises of application of zero mean in each window of the signal that is segmented. This methodology is used as it ensures localized modification of the signal; most biomedical signals have localized respiration noise as observed above. This leads to the exclusion of baseline wandering. Figure 5 indicates the elimination of baseline wandering obtained using the Windowed Zero Mean (WZM) method discussed in this paper.

A bandpass filter [20] is designed using a FIR filter which secures the input cutoff frequencies. The cutoff frequency used for the bandpass here is in the spectrum of 0.5 Hz to 20 Hz. This ensures eradication of the high-frequency noise components present in the signal. Figure 6 displays the result obtained in the spectrum analyzed.

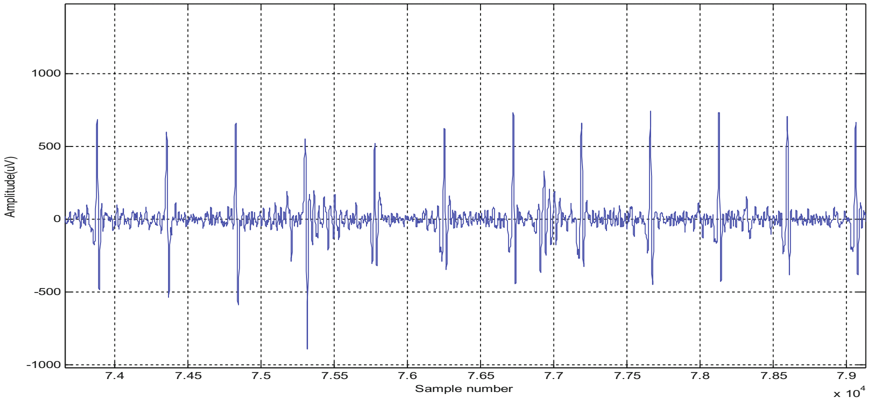


Fig. 6. Windowed Zero Mean (WZM) Signal after passing through bandpass filter

The final signal after application of WZM method and band pass filter [20] is depicted in Fig. 7. It is noted that there is a drastic change in the signal when compared to the original one in Fig. 2.

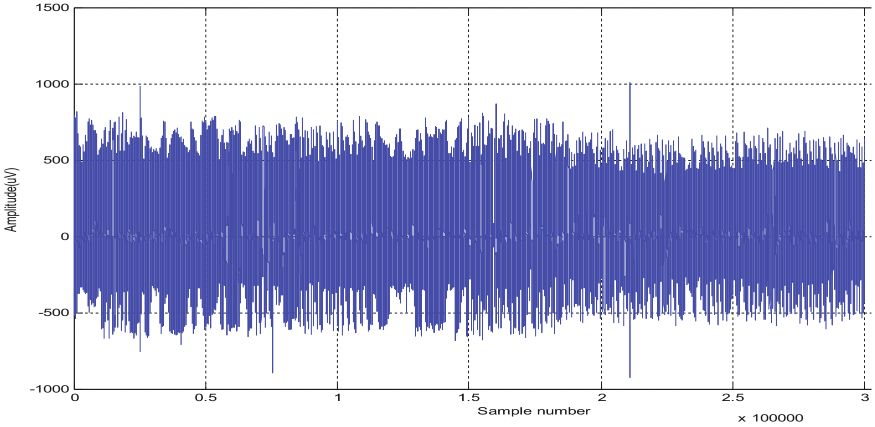


Fig. 7. BW and respiration noise removed abdECG after application of WZM method

Thus, it had been ensured that the elimination of the problems such as baseline wandering and high frequency noise component above was achieved. Now the fECG signal is relatively complemented of such disturbances that can lead to misdiagnosis.

4 Results and Discussion

The Physiobank [17, 18] database used is a renowned database for the disparity in the different types of signals that it possesses. The work on abdomen ECG and validation of result had been done with the help of direct abdECG. The data had been formulated from the signal measured from the abdomen and indicates the problem of baseline wandering and electromyography noise. The data analyzed consists of signals that have a duration of 5 min, a sampling frequency of 1000 Hz and amplitude in microvolts. There are five sets of readings in total with each reading comprising of 5 signal channels out of which the first one in each reading is a directly measured signal from the scalp of the fetus.

On inspection, it had been observed that the range of values for the window size is from 34 to 42. While it could be the sampling frequency of the abdECG signals, the window size for each signal for effective elimination of baseline wander had been individually explored and the same had been tabulated in Table 1.

Table 1 shows the signals tested from the database and their corresponding results which showcase the elimination of the wandering baseline and removal of the high-frequency noise component in a tabulated form.

Some of the cases with anomalies are in case of the fifth signal channel of R08 recording and the first signal channel of R10 recording (which have been marked) where there is irregular variation after application of the windowed zero mean method.

On analysis of the former, it is noted that there is an abrupt peak as denoted in Fig. 8.

Table 1. Results of signal channels analyzed

Recording	R01	R04	R07	R08	R10	Elimination of Baseline wandering	Removal of high-frequency noise component
Signal channel	Window size	Window size	Window size	Window size	Window size		
1	35–37	38–40	39–40	41	34	✓	✓
2	42	39	39–41	36–39	41–42	✓	✓
3	41	39–40	37	42	40	✓	✓
4	39	38	38–39	39	40	✓	✓
5	40–41	38	39	37	35–36	✓	✓

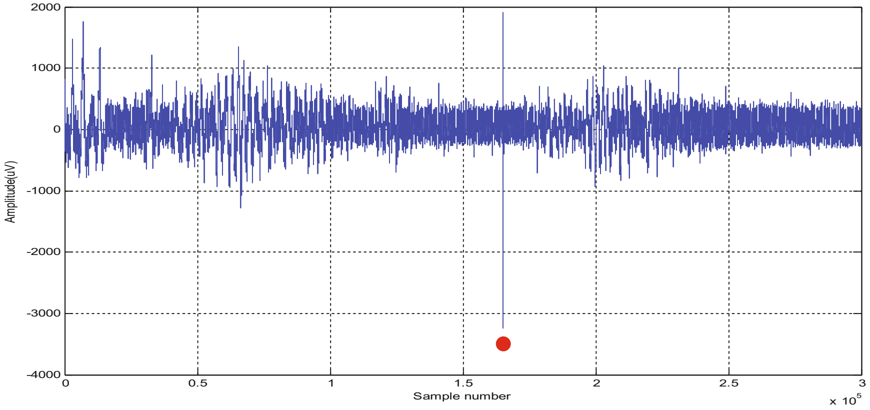


Fig. 8. Original R08 recording Fifth abdECG signal

While it is seen that the ECG signal had been adjusted in Fig. 9 using the WZM method, it had also been observed that it had not been removed entirely; the amplitude, however, had been adjusted, making sure the original state of the ECG signal is not lost.

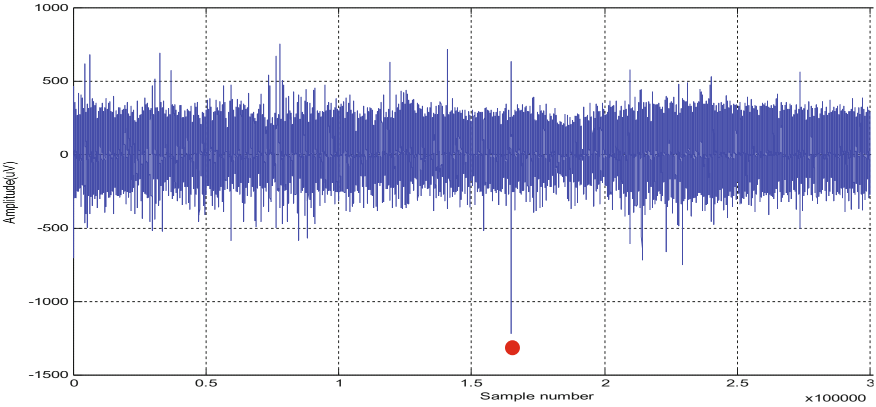


Fig. 9. BW and respiration noise removed R08 recording Fifth abdECG Signal

The second case is the first signal channel of the R10 recording where there is an uneven distribution of the signals as denoted in Fig. 10.

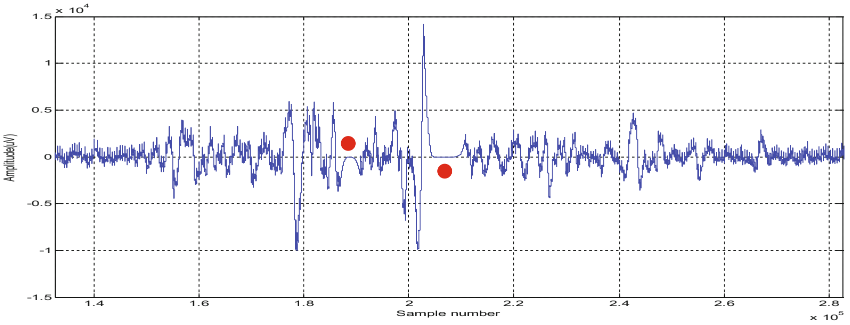


Fig. 10. Original R10 recording First abdECG Signal

While it is noted that that recovery of the signal had been done, there had also been a loss of signal in the filtered signal as shown in Fig. 11 upon application of zero mean.

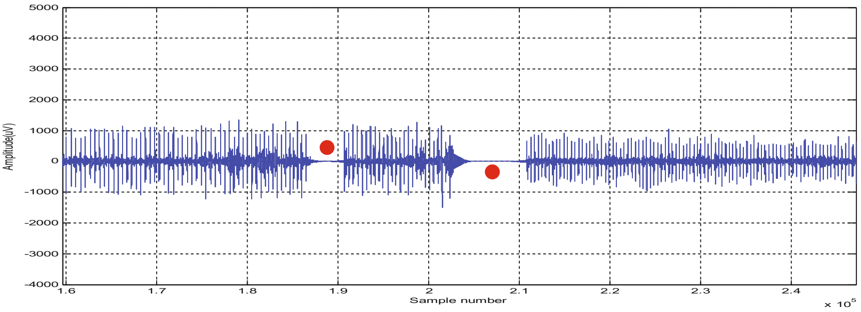


Fig. 11. BW and respiration noise removed R10 recording First abdECG Signal

Hence, it had been discerned that there had been the removal of baseline wandering and high-frequency noise component using the WZM method proposed in this paper.

It is also noted that this method using MATLAB is relatively facile as opposed to other methods as it's time of execution is 3.26 s on an average. The execution time-signal duration ratio is 0.010867. Clinical trials and usage of databases such as MIT-BIH Arrhythmia database can be used for further validation in the future.

5 Conclusion

The high-frequency noise component had been expelled and the problem of baseline wandering had also been addressed. The Windowed Zero Mean method is a much simpler method as opposed to other methods as it does not require a substantial amount of time. Thus, it is a very efficient method as had been proven by assessing the physiobank database. Real-time fECG signals, other signals from the abdomen or parts of the body where such disturbances arise when a recording is taken, as well as other databases can also be used to assess this method. This can help in decipherment and recovery of the ECG signal that is lost as noise. This will aid in the process of fECG measurement for a fetus which can result in proper diagnosis thereby leading to the gratifying welfare of the fetus.

Conflict of interest

The authors disclose that there is no conflict of interest present.

References

1. Mehta, R.S.: ECG and cardiac arrhythmias (2014). https://www.slideshare.net/rsmehta/ecg-arrhythmias
2. UN Inter-agency Group for Child Mortality estimation: Mortality rate, neonatal (per 1,000 livebirths) (2015). http://data.worldbank.org/indicator/SH.DYN.NMRT
3. Kurinczuk, J.J., et al.: The contribution of congenital anomalies to infant mortality. https://www.npeu.ox.ac.uk/downloads/files/infant-mortality/Infant-Mortality-Briefing-Paper-4.pdf
4. Eliasson, H., Sonesson, S.E., Sharland, G., et al.: For the Fetal Working Group of the European Association of Pediatric Cardiology: isolated atrioventricular block in the fetus: a retrospective, multinational, multicenter study of 175 patients, vol. 124, pp. 1919–1926, October 2011. doi:10.1161/CIRCULATIONAHA.111.041970
5. Bar-Cohen, Y., Loeb, G.E., Pruetz, J.D., Silka, M.J., Guerra, C., Vest, A.N., Zhou, L., Chmait, R.H.: Preclinical testing and optimization of a novel fetal micropacemaker. Heart Rhythm (2015). doi:10.1016/j.hrthm.2015.03.022
6. Gini, J.R., Ramachandran, K.I., Nair, R.H., Anand, P.: Portable fetal ECG extractor from abdECG. In: International Conference on Communication and Signal Processing, pp. 0845–0848, April 2016
7. Sarkar, S., Bhattacherjee, S., Pal, S.: Extraction of respiration signal from ECG for respiratory rate estimation. In: Michael Faraday IET International Summit, pp. 336–340, September 2015
8. Park, S.B., Noh, Y.S., Park, S.J., et al.: Med. Bio. Eng. Comput. **46**, 147 (2008). doi:10.1007/s11517-007-0302-y
9. Yacoub, S., Raoof, K.: Noise removal from surface respiratory EMG signal. Int. J. Elect. Comp. Energ. Electron. Commun. Eng. **2**(2), 266–273 (2008)
10. De Luca, C.J., Gilmore, L.D., Kuznetsov, M., Roy, S.H.: Filtering the surface EMG signal: movement artefact and baseline noise contamination. J. Biomech. **43**(8), 1573–1579 (2010)
11. Golabbakhsh, M., Masoumzadeh, M., Sabahi, M.F.: ECG and power line noise removal from respiratory EMG signal using adaptive filters. Majlesi J. Elect. Eng. **5**(4), 28–33 (2011)
12. Blanco-Velasco, M., Weng, B., Barner, K.E.: ECG signal denoising and baseline wander correction based on the empirical mode decomposition. Comput. Biol. Med. **38**, 1–13 (2008)

13. Huang, N.E., Shen, S.S.P.: Introduction to the Hilbert-Huang transform and its related mathematical problems. In: Hilbert-Huang Transform and Its Applications, 2nd edn., pp. 1–11. Abbrev. of Publisher, Singapore (2014). Chap. 1, Sect. 1.2
14. Van Alste, J.A., Schilder, T.S.: Removal of base-line wander and power-line interference from the ECG by an efficient FIR filter with a reduced number of taps. IEEE Trans. Biomed. Eng. **32**(12), 1052–1060 (1985)
15. Pandey, V., Giri, V.K.: High frequency noise removal from ECG using moving average filters. In: International Conference on Emerging Trends in Electrical, Electronics and Sustainable Energy Systems, pp. 191–195, March 2016
16. Zhang, D.: Wavelet approach for ECG baseline wander correction and noise reduction. In: 27th Annual Conference of IEEE Engineering in Medicine and Biology, Shangai, pp. 1212–1215 (2005)
17. Goldberger, A.L., Amaral, L.A.N., Glass, L., Hausdorff, J.M., Ivanov, PCh., Mark, R.G., Mietus, J.E., Moody, G.B., Peng, C.-K., Stanley, H.E.: PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. Circulation. **101**(23), e215–e220 (2000). Circulation Electronic Pages, http://circ.ahajournals.org/content/101/23/e215.full
18. Kotas, M., Jezewski, J., Horoba, L., Matonia, A.: Application of spatio-temporal filtering to fetal electrocardiogram enhancement. Comput. Methods Progr. Biomed. **104**(1), 1–9 (2011)
19. Rakhimov, A.: Normal respiratory frequency, volume, chart. http://www.normalbreathing.com/index-nb.php
20. Kathirvel, P., Sabarimalai Manikandan, M., Prasanna, S.R.M., Soman, K.P.: An efficient R-peak detection based on new nonlinear transformation and first-order gaussian differentiator. Cardiovasc. Eng. Technol. **2**(4), 408–425 (2011)

Early Stage Detection of Diabetic Retinopathy Using an Optimal Feature Set

S.D. Shirbahadurkar[1], Vijay M. Mane[2](✉), and D.V. Jadhav[3]

[1] Department of E&TC Engineering, Zeal COER, Pune, India
[2] Vishwakarma Institute of Technology, Pune, India
vijay.mane@vit.edu
[3] Department of Electronics Engineering, Government Polytechnic,
Ambad, India

Abstract. Diabetic Retinopathy (DR) is the most common source of blindness in the current population worldwide. The development of an automated system will assists to ophthalmologists. DR is a worsening disease, hence early detection is important for diagnosis and proper treatment to prevent blindness. Microaneurysms (MAs) are the first signs of DR; hence their accurate detection is necessary for early stage detection of DR. This paper proposes a three stage system to detect all MAs in the retinal fundus image. First stage extracts all possible candidates using morphological operations and Gabor filter. Feature vector using statistical, gray scale and wavelet features for each candidate is formed in second stage. In the last stage, classification of these candidates as MAs and non MAs is performed using a multilayered feed forward neural network (FFNN) classifier and support vector machine (SVM) classifier. The main objective of the proposed work is to propose a list of important and optimal features for MA detection using the most common features used in the literature. The experiments have been performed on the database DIARETDB1 to evaluate the proposed system. The evaluation parameters accuracy, sensitivity and specificity are obtained as 92%, 79%, 90% and 95%, 76%, 92% respectively for FFNN and SVM classifiers.

Keywords: Diabetic Retinopathy · Microaneurysms · Classifier · Neural network classifier · Gabor filter

1 Introduction

Diabetes is a chronic disease and found in a large number of working populations in most of the developed and developing countries. Diabetes develops when the body does not create sufficient quantity of insulin or fails to process it properly. This increases glucose level in the body which causes damage to almost all organs of the body [1]. The most common damage due to diabetes causes in feet and visual system. The damage to the retina because of diabetes is called diabetic retinopathy (DR).

DR results because of damage to the retinal blood vessels. The blood vessels of the eyes become blocked or swollen due to which they leak the blood into the retina. DR causes damage to the retina without showing any indications at the early stage [2].

S.M. Thampi et al. (eds.), *Advances in Signal Processing and Intelligent Recognition Systems*, Advances in Intelligent Systems and Computing 678,
DOI 10.1007/978-3-319-67934-1_2

Treatment at the later stage is complicated and almost impossible. Hence the before time detection of DR is important to avoid failure of the vision in the patients.

The non-proliferative DR (NPDR) occurs due to blood vessels leak the blood in the retina. Proliferative DR (PDR) is the next stage to NPDR; this causes blindness in the patient. Microaneurysms (MAs) are the before time symptom of DR, which are formed by small swellings near tiny blood vessels. Hemorrhages, hard exudates, and soft exudates are other signs of NPDR. A number of lesions present in the retina decide the stage of NPDR as mild, moderate or severe. The retinal fundus image with signs of DR is shown in Fig. 1.

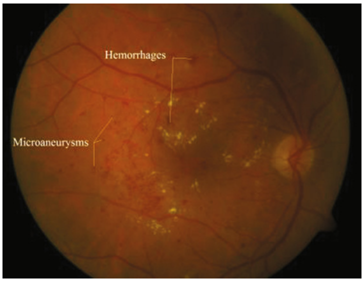


Fig. 1. Retinal fundus image showing early stage signs of DR

For the detection and treatment of DR, regular examination of the retina of the patient is necessary before it affects the sight of the patient. Manual detection is costly, time-consuming and resource demanding. Hence, there is a need to develope an automated system for early stage detection of DR. The system must be sensitive to distinguish images of without DR from images with DR [3].

MAs occur at the end of tiny blood vessels due to swelling of the vessel walls. They are round shaped red color dots of different sizes ranging from 10 to 125 µm. MA detection is a difficult process since they are of the same colour as of blood vessels and variable in sizes. Automatic detection of MAs will lesser the cost as well as difficulties faced in manual detection.

This paper presents a system for automatic detection of MAs for eaarly stage detection of DR. In the proposed system, the image is preprocessed for the removal of noise, shade correction and removal of blood vessels. The main contribution of the proposed system is to show the accuracy of the detection mainly depends on the set of features extracted for the classification. A set of features are porposed for the extracted candidate MAs. The support vector machine and neural network classifiers are used for final classification into MAs and non-MAs. The rest of the paper is organized as, Sect. 2 presents the related work, Sect. 3 elaborates the proposed methodology, Sect. 4 presentes the results obtained and Sect. 5 concludes the paper.

2 Related Work

Quellec et al. [4] proposed a general framework for automated detection of lesions in the fundus image. The feature space was obtained from reference images presentating target lesions using factor analysis. Giancardo et al. [5] used Radon transform to identify lesions using minimum preprocessing and without previous knowledge of retinal morphological features. Antal et al. [6] presented a multi-level ensemble-based approach, where ensemble was formed by a specific group of various preprocessing schemes and candidate extraction schemes. An optimal feature set was used to classify the MAs. Sopharak et al. [7, 8] performed preprocessing, candidate detection using the extended-minima transform, and a Bayesian classification to perform the pixel-level MA detection. Ram et al. [9] presented a two stage classification methodology to detect MAs. The candidate MA lesions were detected using thresholding. The separation of MAs from blood vessels was done using first classifier. The second classifier was applied for final detection of MAs. Haloi et al. [10] detected MAs in color images using deep neural networks. Sinthanayothin et al. [11] detected both MAs and hemorrhages from the fundus images. They enhanced the red lesions using a moat operator. The final set of candidate lesions were extracted after removal of blood vessels. A distinctive method was introduced by Lazar et al. [12, 13] using an unsupervised classification method. The uniqueness of this technique is to distinguish between vessels and MAs by using a 1D scan line at different directions for each pixel. They formed a probability map for each pixel and then by using simple thresholding the final set of candidate's were formed for classification. Akram et al. [14] proposed a system using Gabor filter banks to extract candidate regions. A hybrid classifier combining Gaussian mixture model, SVM and extension of multimodal mediode based modeling approach was used for classification. The study shows that MA detection is a difficult task since they are small in size and are of color same as that of blood vessels. This paper presents an approach to overcome limitations of previous methods by combining different stages. The main objective of the proposed system is to propose a list of important and optimal features for accurate detection of MAs.

3 Proposed Methodology

This paper presents a three stage system for automatic MA detection. The stages of the proposed system includes - extracting the candidate MAs, to form the feature vector for each candidate MA and finally classify them as true MAs or non MAs. The block diagram of the proposed system is as shown in the Fig. 2.

A. Candidate Region Extraction

Pre-processing of the fundus images is required to improve the quality of an input retinal fundus image. Morphological operations such as opening and closing are performed to enhance the MAs in retinal fundus image using Eqs. (1) and (2) respectively, where, *f* is input image and *b* is structuring element. Morphological opening operation is erosion followed by dilation. Morphological closing operation is dilation followed by erosion.

$$f \circ b = (f \ominus b) \oplus b \tag{1}$$

$$f \bullet b = (f \oplus b) \ominus b \tag{2}$$

The green plane of the input color fundus image is selected for further processing. The adaptive histogram equalization is performed to remove noise and brightness variations in the fundus image. The morphological opening is used to smooth the optical disk. Figure 3 shows the outputs of the preprocessing steps.

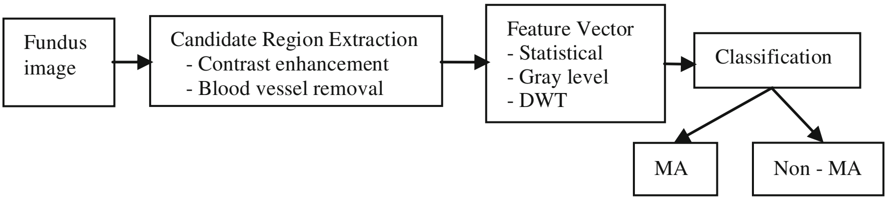


Fig. 2. Block diagram of the proposed system

The Gabor filter is applied to the preprocessed image for blood vessel enhancement. The normalized Gabor filter response is obtained by spanning theta from 0° to 360° at the step of 45°. Gabor filter is characterized by Gaussian kernel function. This interprets different types of shapes based on values of parameters. A 2-D Gabor filter function in the spatial domain is given in Eq. (3), where, f is spatial frequency, θ is orientation angle, γ is aspect ratio and η is wavelength.

$$\begin{aligned} (x, y) &= \frac{f^2}{\pi\gamma\eta} e^{-\left(\frac{f^2}{\gamma^2}x'^2 + \frac{f^2}{\eta^2}y'^2\right)} e^{j2\pi f x'} \\ x' &= x\cos\theta + y\sin\theta \\ y' &= -x\sin\theta + y\cos\theta \end{aligned} \tag{3}$$

The Gabor filter in Eq. (3) is simplified version of general 2-D function derived from 1D Gabor elementary function [15].

The accurate segmentation of blood vessels is important to decrease the occurrence of false MAs and to improve the overall accuracy of the system. The blood vessels are removed by applying a suitable threshold to present all possible candidate regions. The normalized Gabor filter output and all extracted candidates after blood vessel pixels removal is shown in Fig. 4.

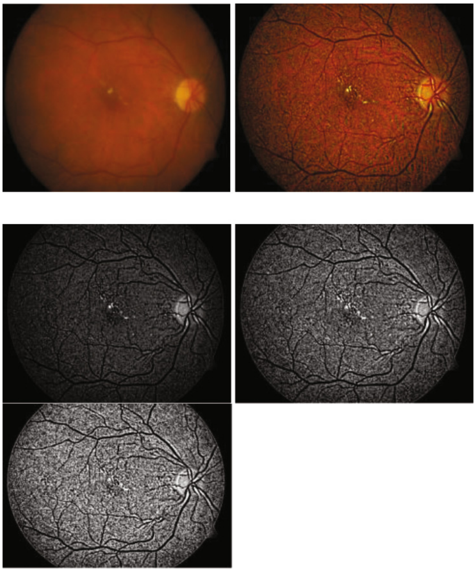


Fig. 3. Pre-processing results (a) Original image (b) Enhanced MAs using morphological operations (c) Green plane (d) Optic disk smoothing (e) Contrast enhanced image (color figure online).

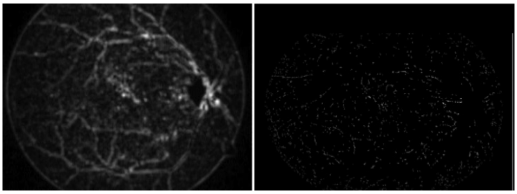


Fig. 4. (a) Normalized Gabor filter output (b) Extracted candidates.

B. Feature Vector Formation

The extracted candidate MAs includes both lesion and non lesion regions. To classify lesions accurately, a proper selection of the features of the regions is very important. A list of features for each candidate lesion is proposed. A feature vector is generated for each extracted candidate MA region to distinguish between MAs and Non-MAs. For an automatic system, selection of optimal features decides the effectiveness of the system. The accurate selection of feature set is very important for the automatic MA detection system. Total twenty eight features are proposed. The feature vector for all candidates is formed using statistical, Gray level Co-occurrence Matrix (GLCM) and wavelet features. Statistical features give shape and area related information of each candidate region. GLCM features are used for extracting information related to intensity of the candidate regions. For the wavelet features, first level DWT transform for 2D images are applied to obtain coefficient matrices for approximation, horizontal details, vertical details and diagonal details sub-bands. Approximation coefficients contain significant information in image while other coefficients contain detailed information. For accurate classification absolute features must be extracted to obtain difference between objects, hence approximation coefficients are used to extract wavelet based features. The set of proposed optimal features are as follows:

1. Area: Entire amount of pixels present in candidate region.
2. Aspect ratio: Ratio of width of bounding box to height of bounding box of candidate region.
3. Perimeter: The distance around the boundary of the candidate region.
4. Eccentricity: Ratio of the distance between foci of the ellipse and length of its major axis.
5. Mean intensity: The mean of all the intensity values in candidate region.
6. Major axis length: Length of the major axis of the ellipse in pixels.
7. Minor axis length: Length of minor axis of ellipse in pixels.
8. Compactness: $C = P^2/A$ where P and A are perimeter and area of candidate region.
9. Equivalent diameter: Diameter of the circle which has same area as the candidate region.
10. Roundness: $r = 4\pi A/P^2$ where A and P are area and perimeter.
11. Skewness: A measure of asymmetry of the data around the sample mean.
12. Orientation: The angle involving x axis and main axis of ellipse.
13. Convex hull: The number of vertex of polygon that contains the region.
14. Convex area: Area of the polygon that contains the candidate region.
15. Euler number: The number of objects minus holes in the region.
16. Mean gradient magnitude of pixels in the candidate region.
17. Mean of all the pixel values in the candidate region.
18. Standard deviation for pixels in the candidate region.
19. Contrast: The contrast in the intensity of pixel and its neighbor over candidate region.
20. Correlation: Measure of how correlated is a pixel to its neighborhood over region.
21. Energy: Sum of squared elements of GLCM for candidate region image.
22. Homogeneity: The nearness of the division of elements of GLCM to its diagonal.
23. Entropy: The entropy of grayscale image of candidate regions.
24. Wavelet energy: Sum of squares of all the elements in coefficient matrices.

$$Energy_{wave} = \sum_{i=1}^{m} \sum_{j=1}^{n} C_{a(i,j)}^2$$

25. Wavelet entropy: This feature is used to characterize the randomness texture of the image

$$Entropy_{wave} = \sum_{i=1}^{m} \sum_{j=1}^{n} C_{a(i,j)} \log C_{a(i,j)}$$

26. Wavelet homogeneity: To compute the closeness of the distribution of wavelet coefficients. $Homogeneity_{wave} = \sum_{i=1}^{m} \sum_{j=1}^{n} \frac{C_{a(i,j)}}{1+|i-j|}$

27. Wavelet correlation: Correlation calculates the gray level correlation of elements in coefficient matrices, where, μ_i,μ_j and σ_i, σ_j are mean and standard deviation respectively. $Correlation_{wave} = \frac{(1-\mu_i)(1-\mu_j)C_{a(i,j)}}{\sigma_i\sigma_j}$
28. Wavelet contrast: Contrast measures local intensity level variation in wavelet decomposed image. $Contrast_{wave} = \sum_{i=1}^{m}\sum_{j=1}^{n}(i-j)^2 C_{a(i,j)}$

A feature vector is formed for each extracted candidate MA using above twenty-eight features.

C. Classification

The feature vector formed for the entire candidate MAs were used for classification. A multi layered feed forward neural network and a support vector machine (SVM) classifiers are used for final classification. A multilayer feed-forward artificial neural network relates of input data to an appropriate output. The error signals are used to calculate the updated weights. The error in the output layer is feed back to previous layer, and weights of these layers are updated. The Levenberg- Marquardt (LM) algorithm for weights updating have been implemented for classification. Levenberg-Marquardt neural network (LMNN) algorithm executes a combined training while updating weights, the algorithm shifts to the steepest descent algorithm. This makes the quadratic approximation using the proper training data. Then it converges to the Gauss Newton algorithm to speeds up the convergence.

Support vectors are the data points that are positioned near to the decision surface (a hyper-plane). In SVM input A is mapped to output C, where a ϵ A is some object and c ϵ C is a class label. Training set ($a1$, $c1$), …… (am, cm) is formed by features of all the candidate MAs. Testing set is formed by features of candidates in test input fundus image. All the candidates were individually classified as MAs or Non MAs.

4 Results and Analysis

The experimental evaluation of the proposed system has been performed on publically available digital fundus image database DIARETDB1 [16]. These images were recorded in Kuopio university hospital, Finland. The database contains 89 color fundus images with 84 images having some mild NPDR signs as MAs of DR and 5 images are normal without DR. These images were recorded with a 50° field-of-view with unfamiliar camera situation. The total 1304 lesions are present in the 89 images. The database is also provided the groundtruth, the locations of the MAs in the retinal fundus image. For the experimentation purpose we used 652 lesions for training and remained 652 lesions are tested using these two classifiers. The performance measures accuracy, sensitivity, and specificity were used to find efficiency of the system. The sensitivity defines the correctly classifying the MA lesions; the specificity is the number to correctly classifying non MA lesions. The true results are the accuracy. True positive TP is defined as correctly classified lesions and false negative FN denotes the incorrectly rejected lesions. The true negative - TN, is the correctly rejected lesions. The incorrectly detected lesion is the false positive- FP. MAs detected using both the classifiers

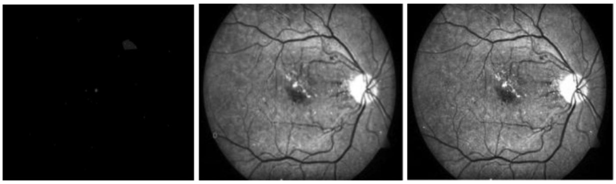


Fig. 5. (a) Ground truth, detected MAs using (b) LMNN (c) SVM

Table 1. Performance measures for two classifiers

Classifier	SVM	LMNN
Factors		
TP	168	182
TN	384	397
FP	52	38
FN	48	35
Precision	.76	.82
Recall	.77	.83
Sensitivity	.77	.83
Specificity	.88	.91
Accuracy	.84	.88

are shown in Fig. 5. Table 1 shows the performance measures for SVM classifier and Levenberg- Marquardt neural network (LMNN) classifier.

The significance of the result is measured by precision. The total number of accurately significant results returned by classifier is a measure called recall. Table 1 shows low false positive rate with high precision values, whereas low false negative rate indicated by high recall values. High values for both the precision and recall show that the classifier yields accurate results.

5 Conclusion

For an automated detection of MAs from retinal fundus image a three stage system is implemented. All possible MAs were extracted and feature vectors of all the candidates along with class labels are given as input to the classifier for training. Classification performance for two classifiers, multilayered feed forward neural network classifier and support vector machine classifier is compared. The proposed system gives better results for detection of individual MAs. The main contribution of the proposed system is the use of an optimal feature set for accurate detection and classification of MAs. In future, the system can be improved for automated screening and to reduce false positive ratios. Also the DR stages can be found using number of lesions present in the retinal fundus image.

References

1. Parenthetic, P., et al.: Diabetic Retinopathy Image Database (DRiDB): a new database for diabetic retinopathy screening programs research. In: 2013 8th International Symposium on Image and Signal Processing and Analysis (ISPA). IEEE (2013)
2. Melville, A., et al.: Complications of diabetes: screening for retinopathy and management of foot ulcers. Qual. Saf. Health Care. **9**(2), 137–141 (2000)
3. Mane, V.M., Jadhav, D.V.: Review: progress towards automated early stage detection of diabetic retinopathy: image analysis systems and potential. J. Med. Biol. Eng. **34**(6), 520–527 (2014)
4. Quellec, G., Russell, S.R., Abramoff, M.D.: Optimal filter framework for automated instantaneous detection of lesions in retinal images. IEEE Trans. Med. Imag. **30**(2), 523–533 (2011)
5. Giancardo, L., Meriaudeau, F., Karnowski, T.P.: Microaneurysm detection with radon transform-based classification on retina images. In: IEEE Engineering in Medicine and Biology Society, pp. 5939–5942 (2011)
6. Antal, B., Hajdu, A.: An ensemble-based system for microaneurysm detection and diabetic retinopathy grading. IEEE Trans. Biomed. Eng. **59**(6), 1720–1726 (2012)
7. Sopharak, A., Uyyanonvara, B., Barman, S.: Simple hybrid method for fine microaneurysm detection from non-dilated diabetic retinopathy retinal images. Comput. Med. Imag. Graph. **37**(5), 394–402 (2013)
8. Sopharak, A., Uyyanonvara, B., Barman, S.: Automatic microaneurysm quantification for diabetic retinopathy screening. In: Proceedings of World Academy of Science, Engineering and Technology, p. 1722 (2013)
9. Ram, K., Joshi, G.D., Sivaswamy, J.: A successive clutter-rejection-based approach for early detection of diabetic retinopathy. IEEE Trans. Biomed. Eng. **58**(3), 664–673 (2011)
10. Haloi, M.: Improved microaneurysm detection using deep neural networks (2015). arXiv preprint arXiv:1505.04424
11. Sinthanayothin, C., et al.: Automated detection of diabetic retinopathy on digital fundus images. Diabet. Med. **19**(2), 105–112 (2002)
12. Lazar, I., Hajdu, A.: Retinal microaneurysm detection through local rotating cross-section profile analysis. IEEE Trans. Med. Imag. **32**(2), 400–407 (2013)
13. Lazar, I., Hajdu, A.: Microaneurysm detection in retinal images using a rotating cross-section based model. In: 2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro, pp. 1405–1409 (2011)
14. Akram, M.U., Khalid, S., Khan, S.A.: Identification and classification of microaneurysms for early detection of diabetic retinopathy. Pattern Recog. **46**(1), 107–116 (2013)
15. Kamarainen, J.K.: Gabor features in image analysis. In: 2012 3rd International Conference on Image Processing Theory, Tools and Applications (IPTA), pp. 13–14 (2012)
16. Kauppi, T., Kalesnykiene, V., Kamarainen, J., Lensu, L., Sorri, I., Raninen, A., Voutilainen, R., Uusitalo, H., Kalviainen, H., Pietila, J.: The DIARETDB1 diabetic retinopathy database and evaluation protocol. In: BMVC, pp. 1–10 (2007)

Exploring Cepstral Coefficient Based Sleep Stage Scoring Method for Single-Channel EEG Signal Using Machine Learning Technique

S. Rajalakshmi[1](✉) and R. Venkatesan[2]

[1] Department of Electrical and Electronics Engineering,
Velammal Engineering College, Surapet, Chennai, India
srajalakshmi312@gmail.com
[2] Department of Electronics and Communication Engineering,
Velammal Engineering College, Surapet, Chennai, India
venky88an@gmail.com

Abstract. Sleep stage scoring is a critical task where conventionally large volume of data has to be analyzed visually which is troublesome, time-consuming and error prone. Eventually, machine learning technique is required for automatic sleep stage scoring. Therefore, a new feature extraction method for EEG analysis and classification is discussed based on the statistical properties of cepstral coefficients. The sleep EEG signal is segmented into 30 s epoch and each epoch is decomposed into different frequency bands: Gamma (γ), Beta (β), Alpha (α), Theta (θ) and Delta (δ) by employing the Discrete Wavelet Transform (DWT). The statistical properties of Mel Frequency Cepstral Coefficients (MFCCs), which represent the short term spectral characteristics of the wavelet coefficients, are extracted. The MFCC feature vectors are incorporated into the Gaussian Mixture Model with Expectation Maximization (GMM-EM) to classify various sleep stages: Wake, Rapid Eye Movement (REM) and Non-Rapid Eye Movement (N-REM) stage1 (S1), N-REM stage2 (S2), N-REM stage3 (S3), N-REM stage4 (S4). The proposed feature extraction for sleep stage scoring achieves 88.71% of average classification accuracy.

Keywords: Cognitive tasks · Discrete Wavelet Transform · Mel Frequency Cepstral Coefficient · Feature extraction · Statistical properties · Gaussian mixture model-expectation maximization

1 Introduction

Sleep scoring is a part of sleep neurobiology closely related to cognitive neuroscience and helps in understanding the neural basis of various cognitive functions such as learning and memory. Therefore sleep stage scoring is of fundamental importance in the neuroscience framework for the discovery of pathologies for instance insomnia, hypersomnia, circadian rhythm disorders, epilepsy and sleep apnea. Consistently, the neuronal system's functional changes are quantified by electrophysiological technique called polysomnography (PSG) involving electroencephalography (EEG), electrooculography (EOG) and electromyography (EMG). Polysomnography is a traditional

S.M. Thampi et al. (eds.), *Advances in Signal Processing and Intelligent Recognition Systems*, Advances in Intelligent Systems and Computing 678,
DOI 10.1007/978-3-319-67934-1_3

standard for sleep stage classification based on visual inspection of physiological signal by sleep specialists according to the Rechtschaffen and Kales's (R&K) guidelines [1] suggested in the year 2012 or the manual suggested by the American Academy of Sleep Medicine in the year 2015 [2]. The R&K rule enables the interpretation of sleep by 30 s epoch into six different stages: Wake, REM, N- REM stages S1, S2, S3 and S4. Whereas, the AASM manual classifies sleep into five stages by combining the N-REM stage 3 and N-REM stage 4 to a single stage. Sleep stage scoring by visual inspection has multitudinous problems: troublesome, time-consuming and fallible due to fatigue. In order to overcome these issues automatic approaches which are sufficiently precise, vigorous, extensible and cost effective have been developed for sleep stage classification. Researches that use multiple physiological signals for automatic sleep stage classification [3] are associated with complex preparation algorithms, limit subject's movements and various problems. In order to overcome the above issues and the EEG signal's capability to study the dynamics of neural information processing of the brain, the automatic sleep stage classification using only the EEG signal gathered the sleep research committee's consciousness. The works [4–7] for automatic sleep stage scoring based on single channel EEG suggests that the single channel EEG based analysis is a suitable way of sleep stage scoring and Pz-Oz channel is more accurate than the Fpz-cz channel. According to result obtained by [7, 8], due to the non-stationary and non-linear characteristics of EEG signal the DWT is much more applicable for sleep stage classification when compared with their counterparts in time domain. Since, the cepstral feature implements framing and windowing of the signal being a part of feature extraction and integrates time-localization information they can be implemented in sleep stage analysis to yield better classification performance. In addition, the cepstral features are more robust in presence of nuisance variation in the signal and finds application in numerous researches [9–11]. This work proposed a new methodology for sleep stage scoring by incorporating the feature extraction method based on the Mel-frequency cepstral coefficients in composite with the DWT. The work is focused on single channel (Pz-Oz) EEG analysis where, the EEG signal is initially segmented into epochs of 30 s duration. Then, the detailed and approximation coefficients are calculated by decomposing the cerebral rhythm into five different sub bands: γ rhythm (>25 Hz), β rhythm (12–25 Hz), α rhythm (6–12 Hz), θ rhythm (3–6 Hz) and δ rhythm (<4 Hz). Each rhythm is associated to specific sleep stage classified by computing the short term power spectrum of the EEG signal based on linear cosine transform of a log power spectrum on a non-linear Mel scale of frequency from the wavelet coefficients. Next, the statistical property of the cepstral coefficient is computed. Finally, the MFCC statistical feature vectors are used to train the GMM-EM classifier for sleep stage classification.

The rest of the paper is arranged as follows: Sect. 2 gives an overview of the existing methods. Section 3 describes the proposed methodology. Section 4 presents the experimental results. Section 5 fetches the conclusion of this work.

2 Related Works

This section discusses the existing methods related to this work. In most researches, the features extracted from the EEG signal are forwarded to the classifier to disintegrate the EEG signal into six possible sleep stages. Suily et al. [12] introduced a new clustering technique for feature extraction and least square support vector machine for classification. The experiment is conducted on the publicly available epileptic EEG, motor imagery EEG data, and mental imagery task EEG dataset with classification accuracy of 94.55%, 84.52% and 61.60% respectively. Bajaj and Pachori [13] suggested smooth pseudo Wigner-Ville distribution to distinguish the EEG signal into different sleep stages based on their Time Frequency images (TFIs). The histograms of the segmented TIFs are used by the multiclass least squares support vector machine for classification. Hsu et al. [14] used the energy features extracted from the EEG signal to differentiate the sleep stages by using the recurrent neural classifier. Herrera et al. [15] employed wavelet transform, Hjorth parameters and symbolic representation to extract different combination of features. The features are ranked using normalized mutual information extraction and fed into SVM classifier for classification. Besides, stacked sequential learning approach is used to improve the classification results.

Sen et al. [16] produced a correlative learning on sleep stage classification by practicing different feature selection: time domain features, frequency domain, time frequency features, linear features and classification algorithms: Random forest, Feed-forward neural network, SVM, radial basis function neural network and decision tree. Zhu et al. [7] introduced sleep stage classification based on single channel EEG by utilizing the concepts of visibility graphs and horizontal visibility graph to extract the features. The corresponding graph features are forwarded to the SVM classifier for classification.

Hafeez allah Amin et al. [17] computed the relative wavelet energy of EEG signal by applying the Discrete Wavelet Transform (DWT). The experimental result shows the comparison of classification performance by SVM, MLP, K-NN and Navie Bayes. Mohammed Diykh et al. [18] suggested a classification method by mapping the derived statistical property and the EEG segment to complex network. K-means classification technique is practiced on two sets of twelve and nine features. Nandini Sengupta et al. [10] proposed a feature set computed from the statistical properties of cepstral coefficients to classify the lung sounds into three different types. It is observed that the statistical property from the cepstral coefficients yield better results when compared to the wavelet coefficients in terms of classification accuracy. It is also observed that the statistical properties from the cepstral coefficients consume less computational overhead in comparison with the baseline cepstral features.

The proposed system introduces cepstral coefficients based feature extraction technique for automatic sleep stage analysis. The proposed system aims to involve extraction of statistical properties from the robust cepstral coefficients. The extracted features are incorporated to GMM-EM pattern recognizer. The classification results are analyzed in terms of different evaluation metrics such as accuracy, sensitivity and specificity. The above discussed metrics are found to be high compared to other conventional methods.

3 Proposed Methodology

The key aspect of this proposed work is to evolve an efficient sleep stage scoring system to classify the single channel EEG into one of the six possible stages according to the R&K recommendation. The sample EEG epoch of each sleep stage is shown in the Fig. 1. The proposed automatic sleep stage scoring predominantly advances through the steps of preprocessing, wavelet decomposition, computation of cepstral coefficient, feature extraction and classification as illustrated in the Fig. 2. In the initial step the sleep EEG signal is segmented into epochs of 30 s duration. In the second step, each epoch is decomposed into different frequency rhythms by applying the DWT. Feature extraction, the third step of the proposed system procures through the compact characterization of large data set without losing distinct information. Then, from the wavelet coefficients the short term power spectrum of the EEG signal based on linear cosine transform of a log power spectrum on a non-linear Mel scale of frequency is computed. The final classification step assigns each epoch to their analogous sleep stage. The MFCC statistical feature vectors are used to train the GMM-EM classifier for sleep stage classification. The test procedure was realized by using the validation data prepared from the 3 hrs data of the subjects SC4001E0 to SC4051E0 of Physionet Bank Expanded sleep EDF database shown in Table 1.

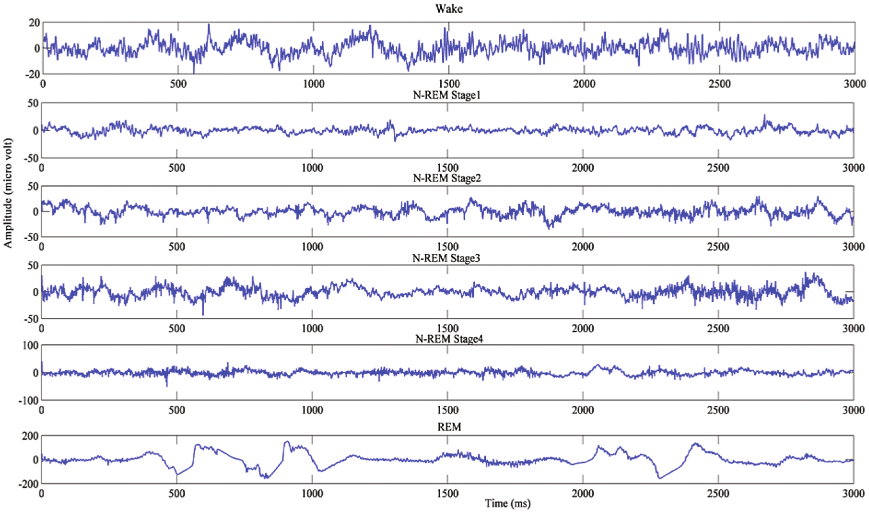


Fig. 1. Sample EEG epoch of various sleep stages

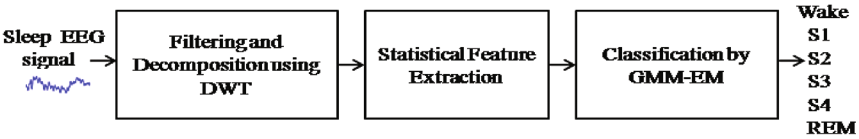

Sleep EEG signal
Filtering and Decomposition using DWT
Statistical Feature Extraction
Classification by GMM-EM
Wake
S1
S2
S3
S4
REM

Fig. 2. Structure of the proposed methodology

Table 1. Validation data prepared for the test procedure

	Wake	Stage 1	Stage 2	Stage 3	Stage 4	REM
Epochs	593	454	1612	507	326	347

3.1 Data Description

The data utilized by the proposed work to conduct the experiment is from the publicly available Physionet's Sleep-EDF data set [19], widely adopted in the literature [6, 7]. The signals are recorded from Caucasian males and females of age ranged from 21–35 years by employing a miniature telemetry system. The signals are recorded at 100 Hz sampling rate during 24 h of subject's daily life. This work utilized the EEG signal from Pz-Oz channel where all others signals are discarded.

3.2 Discrete Wavelet Transform

The DWT is used to extract local features from the biomedical signals especially for EEG signal due to its non-stationary and non-linear characteristics. The concept of DWT is to decompose the signal into multilevel successive frequency rhythms by employing a set of scaling function (ϕ) and wavelet function ψ given by,

$$\mathrm{DWT}\ (\mathrm{j},\,\mathrm{k}) = \frac{1}{\sqrt{|2^{\mathrm{j}}|}} \int_{-\infty}^{+\infty} \mathrm{x}(\mathrm{t})\Psi\left(\frac{\mathrm{t}-2^{\mathrm{j}}\mathrm{k}}{2^{\mathrm{j}}}\right)\mathrm{d}(\mathrm{t}) \tag{1}$$

This represents the signal as a series of approximation coefficient and the detailed coefficient. Where, the approximation coefficient is the outcome of the high pass filter g (n), the discrete mother wavelet and the detailed coefficients is the outcome of the low pass filter h(n), its mirror version. The approximation and the detailed coefficient at the first level decomposition are represented by A1 and D1 respectively. A1 is further disintegrated and the procedure is repeated till the specified number of decomposition level is achieved as shown in the Fig. 3(a). At each level, filtering doubles the frequency resolution and down sampling halves the time resolution. In this work, the normalized DWT of the Daubechies family with two vanishing moment is employed to analyze and decompose the EEG signal into multilevel successive frequency bands. The Db2 is chosen due to its efficiency in capturing the data variation only with two null moments.

The dataset used for this experiment is recorded at a sampling frequency of 100 Hz. Therefore, on obeying the nyquist theorem four levels of decomposition is required to achieve the required frequency bands of the sleep EEG signal. The structure of 4 level DWT adopted in this work with the corresponding frequency range is shown in the Fig. 3(b). Each sleep stage is characterized with particular EEG rhythm tabulated in the Table 2. The four level decomposition of an epoch during which the subject is in REM state is shown in the Fig. 4.

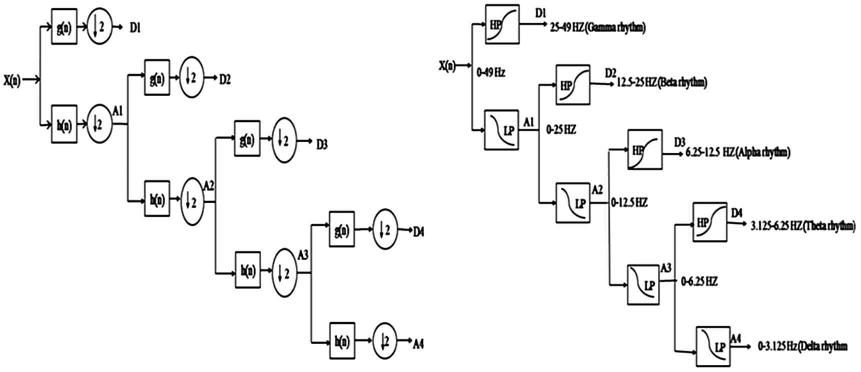


Fig. 3. (a) Structure of 4 levels DWT decomposition (b) DWT decomposition structure of the proposed method

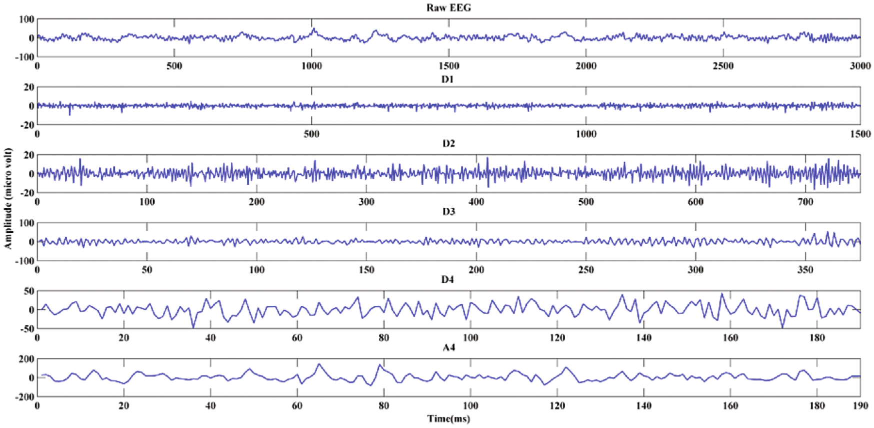


Fig. 4. Wavelet decomposition of an epoch at which the subject is at REM stage

Table 2. EEG rhythms corresponding to the sleep stages

Stage	Name	Rhythm
Wake	Relaxed	Alpha, Beta
Sage 1	Drowsiness	Alpha, Theta
Stage 2	Light Sleep	Theta
Stage 3	Deep Sleep	Delta, Theta
Stage 4	Deep Sleep	Delta
REM	Dreaming	Beta

3.3 Mel Frequency Cepstral Coefficient

Mel Frequency Cepstral coefficient is a static feature extraction method that depends on the spectral analysis of the signal with a fixed resolution along a subjective frequency scale called the Mel Frequency scale. The structure of MFCC feature extraction is shown in the Fig. 5. The input Sleep EEG is firstly framed and windowed. Windowing is a point wise multiplication of the frame and the window function in time domain. The concept of applying the window function is to minimize the spectral distortion to increase the continuity of the adjacent frames. Then, the FFT is applied on the frame and the magnitude of the resulting spectrum is warped onto the Mel-scale. The idea behind the FFT is to represent a signal as the sum of properly chosen sinusoidal waves. The Fast Fourier Transform converts the frames in time domain to frequency domain which is defined on the set of N samples as follows:

$$w(n) = 0.54 - 0.46\cos\left(\frac{2\pi n}{N-1}\right).\ 0 \le n \le N-1 \tag{2}$$

The result after this step is the spectrum or periodogram of the EEG signal. Then the log amplitude of the spectrum is mapped onto the Mel scale using triangular filters to obtain the Mel spectrum. The Mel scale is a mapping between the real frequency scale (Hz) and the perceived frequency scale (Mels). The mapping is virtually linear given by,

$$m = 2595\log_{10}\left(\frac{f}{700} + 1\right) \tag{3}$$

In the next step, the log Mel spectrum is converted back to time domain by applying the Discrete Cosine transform. The final resulting is the Mel Frequency Cepstral Coefficients obtained by,

$$c_n = \sum_{k=1}^{k} \log(S_k)\cos\left[n\left(k - \frac{1}{2}\right)\frac{\pi}{k}\right], n = 0, 1, \ldots k - 1 \tag{4}$$

Where n = 0, 1, …, N − 1, k is the number of filters, N is the number of coefficients and c(n) is the Mel frequency Cepstal Coefficient.

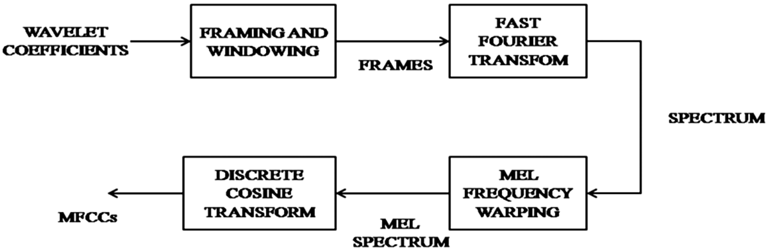


Fig. 5. Structure of MFCC computation

The attractive features of MFCC are confirmed by the histograms of MFCC for various sleep stages illustrated in the Fig. 6. It is noticed that the shape and the range of frequency values are markedly distinct among different stages of sleep conforming the usefulness of MFCC for sleep EEG signal analysis.

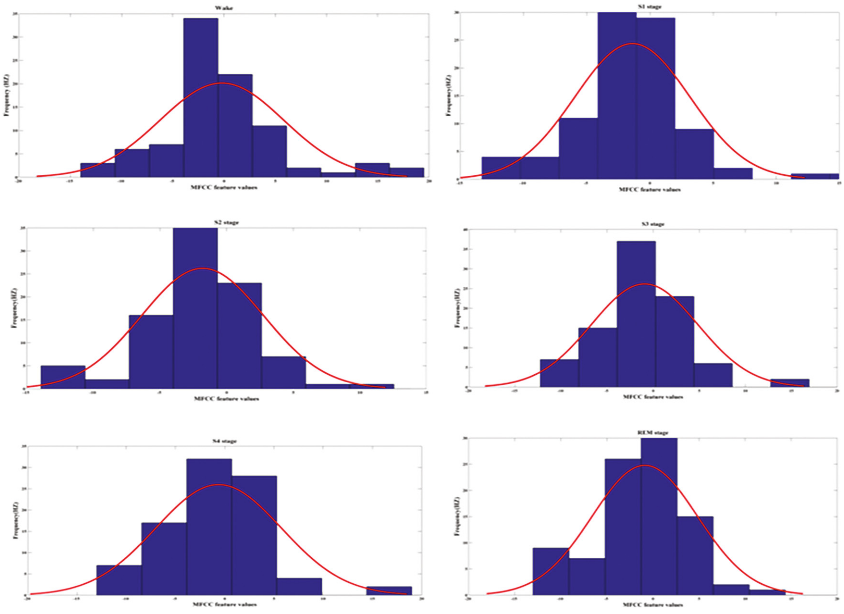


Fig. 6. Histograms of MFCCs for various sleep stages

3.4 Feature Extraction

The feature extraction is used to extract relevant information from the EEG recording for evaluation and understanding of the desired cognitive processes. The main goal of Feature extraction is to reduce the dimensionality of large volume of signal data without any loss of information. The extracted feature has direct impact on the systems classification performance. Hence, extracting suitable features from EEG signals to get high classification performance is mandatory. In this work the statistical features energy, envelope kurtosis, envelope skewness, Standard Deviation and variance are extracted from the MFFCs computed on the wavelet coefficients of each decomposition level.

3.4.1 Energy

The energy of the signal in discrete form is calculated by the given equation,

$$E = T \sum_{n-0}^{N-1} x^2[n] \tag{5}$$

Where, T is the duration and x[n] is the discrete samples.

3.4.2 Variance

Variance is the measure of how far a set of numbers is spread out from its mean.

$$\sigma^2 = \frac{\sum (x - \mu)^2}{N} \tag{6}$$

Where, x is variable, N is the number of variables and μ is the mean.

3.4.3 Standard Deviation

Standard deviation is the measure of dispersion of a dataset given by,

$$\mathrm{Std}(x) = \sqrt{\frac{\sum (x - \mu)^2}{N}} \tag{7}$$

Where, x is the variable, N is the number of variables and μ is the mean.

3.4.4 Envelope

Envelope is a smooth curve outlining the extremes of the signal detected by using the Hilbert transform defined by,

$$h_x = \text{hilbert transform}(x) \tag{8}$$

Where, x is the input dataset.

3.4.5 Envelope Kurtosis

Kurtosis is a non-dimensional quantity that measures the peakedness of a dataset.

$$\mathrm{Kurt}\left(\hat{f}(x)\right) = \frac{E\left(\hat{f}(x) - \mu_{\hat{f}(x)}\right)^4}{\left(E\left(\hat{f}(x) - \mu_{\hat{f}(x)}\right)^2\right)^4} \tag{9}$$

Where, h_x is the Hilbert transform of the dataset x and μ_{hx} is the mean of h_x

3.4.6 Envelope Skewness

Skewness defines the extent to which a distribution differs from a normal distribution.

$$\mathrm{Skew}\ \left(\hat{f}(x)\right) = \frac{E\left[\left(\hat{f}(x) - \mu_{\hat{f}(x)}\right)^3\right]}{\left(E\left[\left(\hat{f}(x) - \mu_{\hat{f}(x)}\right)^2\right]\right)^{\frac{3}{2}}} \tag{10}$$

Where, h_x is the Hilbert transform of the dataset x and μ_{hx} is the Mean.

3.5 Classification

To reveal the productiveness of the proposed feature extraction scheme in cognitive function classification, the GMM classifier is used.

3.5.1 Gaussian Mixture Model-Expectation Maximization

In GMM based model the random variable y is represented as a weighted sum of G number of Gaussian functions which are widely used for automatic identification of bio signals [20, 21] and for the approximation of continuous probability density function from a multi-dimensional feature. The multivariate Gaussian probability density given by,

$$p(y) = \sum_{\mathrm{j}=1}^{G} q_j \mathrm{N}(\mathrm{y}, \mu_{\mathrm{j},} \sigma_j) \tag{11}$$

Where $\mathrm{N}(\mathrm{y}, \mu_{\mathrm{j},} \sigma_j)$ is the n dimensional data vector of normal distribution with covariance σ_j and mean $\mu_{\mathrm{j},}$. The q_i is the weight representing the probability of class j defined as

$$N(\mathrm{y}, \mu_{\mathrm{j},} \sigma_j) = \frac{1}{\sqrt{(2\pi)^{\mathrm{n}} |\sigma_j|}} \exp\left(\frac{-1}{2} (\mathrm{y} - \mu_j)^T \sigma_{\mathrm{j}}^{-1} (\mathrm{y} - \mu_j) \right) \tag{12}$$

According to Bayes' rule the conditional probability of the observation vector s belongs to the component g_i of the GMM defined by,

$$p\left(\frac{g_j}{s}\right) = \frac{q_j \mathrm{N}(\mathrm{y}, \mu_j, \sigma_j)}{\sum_{\mathrm{j}=1}^{G} q_k \mathrm{N}(y_k, \mu_k, \sigma_k)} \tag{13}$$

The Expectation-Maximization (EM) procedure is used to approximate the q, μ and σ variable that yields the maximum likelihood of the observed data D. The parameters of the mapping function are obtained by the joint probabilistic density of source and target features. A joint feature vector $Y = [s^T, t^T]^T$ where, s and t are the time aligned input and output feature vectors which are utilized to evaluate the GMM variables. The mapping function is defined by

$$M(y) = \sum_{\mathrm{z}=1}^{G} \frac{g_z}{s} \left[\mu_z^t \sigma_z^{\mathrm{ts}} \sigma_z^{\mathrm{ss}^{-1}} (\mathrm{s} - \mu_z)^s \right] \tag{14}$$

Where, μ_z is the mean and σ_z is the covariance of rth Gaussian distribution. In GMM based technique the number of Gaussian functions G is determined by the amount of training sample.

4 Result and Discussion

The EEG signal recorded during the sleep provides useful information regarding the sleep stages which are useful in the diagnosis of sleep related disorders namely epilepsy, depression, sleep apnea and stress diagnosis. This work attempt to solve the problem of conventional sleep stage classification by employing the statistical features of cepstral coefficient computed from the DWT sub-bands of single channel Sleep EEG signal. The confusion matrix is constructed between the developed work adopting the GMM-EM classifier and the Experts' method of visual scoring using R&k manual and is shown in the Table 3.

Table 3. Confusion matrix between the scoring result by GMM-EM classifier and experts scoring

Expert Scoring	Scoring result by GMM-EM classifier						
	Wake	Stage 1	Stage 2	Stage 3	Stage 4	REM	Success Rate in %
Wake	**470**	74	17	8	4	20	79.25
S1	84	**187**	90	7	5	81	41.18
S2	60	111	**1145**	87	89	120	71.08
S3	12	17	20	**247**	207	120	48.71
S4	7	9	13	139	**148**	4	45.4
REM	10	15	28	3	1	**290**	83.57

Table 4. Performance evaluation of the proposed work using GMM-EM classifier

Sleep Stages	Accuracy	Precision	Sensitivity	Specificity
Wake	92.8	73.1	79.25	94.7
S1	87.2	45.27	41.189	93.32
S2	83.5	87.20	71	92.46
S3	84.9	50.38	48.72	92.68
S4	87.4	32.6	45.4	91.3
REM	95	68.23	83.57	96.13

The stage S1 is a transition phase and is a combination of wakefulness and sleep resulting in similarity with the neuronal oscillations of S1 and wake. Therefore, the classification of S1 stage is an extensive challenge to any sleep stage scoring system. Due to which S1 is misclassified as wake or REM. The proposed system classifies 42.18% of S1 epochs correctly. The REM stage is of particular importance which accounts for 5–20% of whole night sleep necessary for the diagnosis of various sleep related disorders including REM behavior disorder (RBD), narcolepsy etc. The proposed method detects 83.57% of REM epochs correctly. Also, the proposed method classifies 79.25%, 71.03%, 48.71% and 45.4% of epochs as Wake, S2, S3 and S4 respectively. In addition, the performance of the developed system is determined by evaluating the parameters: precision, accuracy, sensitivity and specificity shown in

Table 4. This work achieves an average of 59.46% of precision, 88.72% of accuracy, 61.52% of sensitivity and 93.43% of specificity.

The duration of various sleep stage is not even and therefore the recorded EEG signal has uneven number of epochs for various stage which can be confirmed by an observation on the Table 1. Thus, the automatic sleep stage scoring system poses expansive challenge due to class imbalance problem where in the specimen belonging to a class in a training data-set are higher in number than the specimen belonging to other class. Therefore, the classification model is liable to favor classifying all the samples belonging to the majority class. Even though the proposed method performs better, the class imbalance problem stops it from achieving 100% accuracy in all the cases of interest. As, the cepstral feature MFCC accomplish framing and windowing of the signal they integrates time-localization information which is averse to the wavelet-based method. Thus proposed MFCC based statistical features for sleep stage scoring provides an average classification accuracy of 89.71%. Since, the work is implemented on single channel EEG the method does not require any filtering, artifact rejection, and noise removal algorithm. This is a major advantage of the proposed methodology which can be convenient for portable sleep quality evaluation devices. As, the sleep stage scoring scheme can be operated directly on the recorded signal the device will ensure reduced power consumption.

The classification results of wavelet-based features are demonstrated in [22]. On comparing the classification performance of the wavelet-based features and cepstral features, it is found that the cepstral features performs 0.99 times better than the wavelet-based features. This may be due to the reason that the short-term spectral characteristics of the EEG signal can represent the sleep stage information in a more effective manner unlike the statistical measures of EEG signal in wavelet domain.

5 Conclusion

The paper discusses the exploring of Mel Frequency Cepstral coefficients for automatic sleep stage scoring based on single-channel EEG. The current study exploits the statistical parameters of the Mel Frequency Cepstral Coefficients for the purpose of classifying sleep stages. The proposed automatic sleep stage scoring system provides more accurate and speedy diagnosis of EEG signals corresponding to complex cognitive tasks and it will be useful in the clinical application such as epilepsy, depression, sleep apnea and stress diagnosis. MFCC based feature extraction technique achieve better performance than the wavelet based features and relatively robust for sleep stage classification. Since the proposed framework is capable of analyzing the non-stationary EEG signal it can also be implemented in real time EEG based signal analysis including Brain-Computer Interface(BCI) investigation to control external devices using cognitive neuroscience.

References

1. Rechtschaffen, A., Kales, A.: Manual of standardized terminology, techniques and scoring systems for sleep stages of human subjects. U.G.P. Office, Washington DC Public Health Service (2012)
2. Iber, A.L.C.C., Ancoli-Israel, S., Quan, S.F.: The AASM manual for the scoring of sleep and associated events: rules, terminology and technical specification. American Academy of Sleep Medicine, Westchester, USA (2015)
3. Lajnef, T., et al.: Learning machines and sleeping brains: automatic sleep stage classification using decision-tree multi-class support vector machines. J. Neurosci. Methods. **250**, 94–105 (2015)
4. Hsu, L., Yang, T., Wang, J., Hsu, C.: Automatic sleep stage recurrent neu-ral classifier using energy features of EEG signals. Neurocomputing **104**, 105–114 (2013)
5. Ronzhina, M., Janousek, O., Kolarova, J., et al.: Sleep scoring using artificial neural networks. Sleep Med. Rev. **16**, 251–263 (2012)
6. Zhu, G., Li, Y., Wen, P.: Analysis and classification of sleep stages based on ifference visibility graphs from a single-channel EEG signal. IEEE J. Biomed. Health Inform. **18**(6), 1813–1821 (2014)
7. Single-channel EEG sleep stage classification based on a streamlined set of statistical features in wavelet domain. Med. Biol. Eng. Comp. (2017). doi:10.1007/s11517-016-1519-4
8. Subasi, A.: Automatic recognition of alertness level from EEG by using neural network and wavelet coefficients. Expert Syst. Appl. **28**(4), 701–711 (2005)
9. Sengupta, N., Sahidullah, M., Saha, G.: Lung sound classification using cepstral-based statistical features. Comp. Biol. Med. (2016). doi:10.1016/j.compbiomed.2016.05.013
10. Biagetti, G., et al.: Speaker identification with short sequences of speech frames. In: Proceedings of the International Conference on Pattern Recognition Applications and Methods, vol. 2. SCITEPRESS-Science and Technology Publications, Lda (2015)
11. Hokking, R., Woraratpanya, K., Kuroki, Y.: Speech recognition of different sampling rates using fractal code descriptor. In: 13th International Joint Conference on Computer Science and Software Engineering (JCSSE). IEEE (2016)
12. Li, S.Y., Wen, P.P.: Clustering technique-based least square support vector machine for EEG signal classification. Comp. Methods Prog. Biomed. **104**, 358–372 (2011)
13. Bajaj, V., Pachori, R.B.: Automatic classification of sleep stages based on the time-frequency image of EEG signals. Comput. Methods Programs Biomed. **112**, 320–328 (2013)
14. Hsu, Y.-L., Yang, Y.-T., Wang, J.-S., Hsu, C.-Y.: Automatic sleep stage recurrent neural classifier using energy features of EEG signals. Neurocomputing **104**, 105–114 (2013)
15. Herrera, L.J., Fernandes, C.M., Mora, A.M., Migotina, D., Largo, R., Guillén, A., Rosa, A. C.: Combination of heterogeneous EEG feature extraction methods and stacked sequential learning for sleep stage classification. Int. J. Neural Syst. **23**, 1350012 (2013)
16. Şen, B., Peker, M., Çavuşoglu, A., Çelebi, F.V.: A comparative study on classification of sleep stage based on EEG signals using feature selection and classification algorithms. J. Med. Syst. **38**, 1–21 (2014)
17. Amin, H.U., Malik, A.S., Ahmad, R.F., Badruddin, N., Kamel, N., Hussain, M., Chooi, W. T.: Feature extraction and classification for EEG signals using wavelet transform and machine learning techniques. Australas. Phys. Eng. Sci. Med. (2015). doi:10.1007/s13246-015-0333-x
18. Diykh, Mohammed, Li, Yan: Complex networks approach for EEG signal sleep stages classification. Expert Syst. Appl. (2016). doi:10.1016/j.eswa.2016.07.004

19. Kemp, B., Olivan, J.: European data format ‘plus’ (EDF+), an EDF alike standard format for the exchange of physiological data. Clin. Neurophysiol. **114**, 1755–1761 (2003)
20. Spadaccini, A., Beritelli, F.: Human identity verification based on heart sounds: recent advances and future directions. In: Biometrics, pp. 217–234. InTech (2011)
21. Frean, M., Lilley, M., Boyle, P.: Implementing Gaussian process inference with neural networks. Int. J. Neural Syst. **16**(5), 321–327 (2006)
22. Rajalakshmi, S., Prakash, R., Venkatesan, R., Balaji Ganesh, A.: Sleep stage scoring based on single-channel EEG using machine learning technique. In: International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS) (2017)

Non Linear Tracking Using Unscented Kalman Filter

P. Sudheesh(✉) and M. Jayakumar

Department of Electronics and Communication,
Amrita School of Engineering, Coimbatore, Amrita Vishwa Vidyapeetham,
Amrita University, Coimbatore, India
{p_sudheesh,m_jayakumar}@cb.amrita.edu

Abstract. Accurate localization of mobile robots to locate its position and orientation is of key importance since it enables a mobile robot to navigate properly in any given environment. Various techniques of localization used are such as GPS/GNSS, IMU sensors or by using odometric measurements. However each of these techniques suffers from various drawbacks. Dead-reckoning (DR) is a popular client to get precise localization information. DR estimates the current position based on the previous positions observed over a span of time. However DR depends on encoder and odometric information which are subject to major errors due to surface roughness, wheel slippage and tolerance rate of the machine which leads to an accumulation of errors. Many researchers have addressed this problem by adding certain external sources such as encoded magnetic compass, rate-gyros etc., However addition of these sensors has led to various new errors. In this paper, the use of unscented Kalman filter (UKF) is proposed along with the DR to get accurate localization information. UKF uses a deterministic sampling approach that captures the estimates of mean and covariance with a set of sigma points. The simulation results show that the proposed method is able to track the desired path with least error when compared to DR used alone. The localization of a mobile robot with the proposed system is also highly reliable.

Keywords: Non linear tracking · Mobile robots · Unscented kalman filter

1 Introduction

Navigation of autonomous robots in any environment depends on accurate localization information. There are many techniques available for the localization mobile robots. The most widely used technique among them is the dead reckoning method [10]. In cases where GPS/GNSS is used, precise localization is achieved by receiving signals from three to four satellites simultaneously. However, there are many situations in which GPS/GNSS cannot be used, for example, indoor, underwater [13], or extra-terrestrial. In case of urban environments, very strong multipath propagation occurs and thus the signal quality is degraded one way or the other.

So the dead-reckoning method is a popular candidate overcoming these limitations [4]. It provides reliable information by fusing information from various sensors and

S.M. Thampi et al. (eds.), *Advances in Signal Processing and Intelligent Recognition Systems*, Advances in Intelligent Systems and Computing 678,
DOI 10.1007/978-3-319-67934-1_4

calculates current position with reference to an inertial frame of reference based on previously determined position.

Many researchers have worked on the mobile robot localization problem. In most of the techniques used, the researchers have added some external sources to overcome the problems in the localization of the robot. Kim and Seong [6] used a location system that uses encoded magnetic compass which accounts for the drift due to slippage of wheel. However, it didn't function well in places where the magnetic fields keep on varying. Song and Seun [9] used a low-cost rate-gyro which overcomes the problem of slippage of wheels. However, with the use of low cost rate-gyro, there is a problem of drift rate getting higher. Hence, introducing a new sensor in itself introduces errors [1].

The dead-reckoning method depends on encoder or odometric information [7] which is subject to major errors due to wheel slippage, tolerance rate of machine and surface roughness. Hence there is an accumulation of errors which affects the localization [3]. This problem can be solved by using Kalman filter (KF) with the DR to get reliable localization information [2]. KF is a mathematical tool that uses a set of equations recursively to estimate the future positions effectively. The KF mainly deals with the uncertainties in modelling of the system. The KF is mainly used for linear systems. EKF and UKF are the enhanced versions of KF that deals with non-linear systems. The EKF linearizes all the non-linear models. But there are two important drawbacks in case of EKF. First, the Jacobian matrices derivation can be complex making the implementation difficult [8]. Second, if intervals for linearization are not sufficient, it may lead to filter instability [12]. The UKF approximates a Gaussian distribution rather than a non-linear function and uses a deterministic sampling approach that captures the estimates of mean and covariance with a set of sigma points. Thus, UKF performs better than the EKF for non-linear systems [11]. In this paper, the mobile robot localization problem is approached by using UKF with the DR.

2 System Model

The system proposed here is applied for wheeled mobile robot (WMR) WMR consists of three wheels, of which two wheels in rear of the chassis with the same axis acts as driving wheel with each wheel driven independently with motor and the third wheel at the front of the chassis is a free wheel. The non-linear equation given in Eq. (1) describes the kinematic model of the WMR,

$$\begin{bmatrix} x_{st} = v_r . \cos(\theta_{st}) \\ y_{st} = v_r . \sin(\theta_{st}) \\ \theta_{st} = \omega_\Upsilon \end{bmatrix} \tag{1}$$

where the position of the mobile robot is given by the x_{st} – and y_{st} – coordinates, the angle between the X-axis and position direction is given by θ_{st}, that is, orientation of the robot, linear velocity is given by v_r, and the angular velocity is given by ω_r.

2.1 Dead Reckoning

The dead reckoning method calculates the current orientation and position of the moving WMR based on the previous information on orientation and position. At time, $t = t_{k+1}$, the current position and orientation of the robot are calculated based on the previous readings of orientation and position at time $t = t_k$, By using Euler's approximation Eq. (1) becomes

$$\begin{bmatrix} x_{st}(k+1) = x_{st(k)} + v_{r(k)Ts}\cos\left(\theta_{st(k)}\right) \\ y_{st(k+1)} = y_{st(k)} + v_{r(k)Ts}\sin\left(\theta_{st(k)}\right) \\ \theta_{st(k+1)} = \theta_{st(k)} + \omega_{r(k)Ts} \end{bmatrix} \tag{2}$$

where $T_s = t_{k+1} - t_k$ is the sampling period [5]. The WMR consists of an encoder which releases pulses and are used to calculate the distance travelled by the robot with the use of DC motor attached to each wheel. Equation (3) given below gives the mathematical model

$$\begin{bmatrix} x_{st(k+1)} = x_{st(k)} + \dfrac{\pi D_w}{2}\dfrac{(\Delta T_L + \Delta T_R)}{T_w}\cos\left(\theta_{st(k)}\right) \\ y_{st(k+1)} = y_{st(k)} + \dfrac{\pi D}{2}\dfrac{(\Delta T_L + \Delta T_R)}{T_w}\sin\left(\theta_{st(k)}\right) \\ \theta_{st(k+1)} = \theta_{st(k)} + \dfrac{\pi D_w}{dst}\dfrac{(\Delta T_R - \Delta T_L)}{T_w} \end{bmatrix} \tag{3}$$

where $\Delta T = T_{k+1} - T_k$ represents impulse of the encoder (the left wheel encoder is represented by ΔT_L and the right wheel encoder is represented by ΔT_R), T_s is the sampling period, dst is the difference in length in between the wheels, D_w represents the diameter of the wheel and T_w represents the number of encoder pulses for complete rotation.

3 Unscented Kalman Filter Based Tracking

The Unscented Transform is a non-linear transform which converts the state vector st_k to a set of points called sigma points, χ_{k-1} based on apriori conditions. st_k is a random variable which is assumed to be Gaussian distributed of length $L \times 1$ with a mean $\widetilde{st_k}$ and covariance matrix P_x The length of χ_k is $2L \times 1$.The scaling parameters λ (composite scaling parameter), α (primary scaling parameter), β (secondary scaling parameter) and κ (tertiary scaling parameter) affect the sigma points spread and also determine the weight vectors which help in re-building the aposteriori statistics. R and Q respectively are the measurement and process noise covariance. $w^{(c)}$ and $w^{(m)}$ are the weights. The steps involved in the algorithm is given by

1. Determining weights and scaling parameters and initializing Q and R.

$$\lambda = \alpha^2(L+K) - L \tag{4}$$

$$W_o^{(m)} = \frac{\lambda}{L+\lambda} \tag{5}$$

$$W_o^{(c)} = \frac{\lambda}{L+\lambda} + 1 + \beta - \alpha^2 \tag{6}$$

$$W_i^{(m)} = W_i^{(c)} = \frac{\lambda}{2(L+\lambda)}, \; i = 1, \ldots, 2L \tag{7}$$

2. Calculating square-root of p_{k-1} using Cholesky Decomposition as shown in Eq. (8) and calculating the sigma-points using Eq. (9)

$$\sqrt{p_{k-1}} = chol\,(p_{k-1}) \tag{8}$$

$$k_{k-1} = \begin{bmatrix} st_{k-1} + \sqrt[st_{k-1}]{(\lambda+1) \times p_{k-1}} \\ st_{k-1} - \sqrt{(\lambda+1) \times p_{k-1}} \end{bmatrix}^T \tag{9}$$

3. Prediction Transformation
   a. Each sigma-point is propagated through a non-linear function $f()$

$$\chi(k|k-1) = f(\chi_{k-1}) \tag{10}$$

b. With the transformed sigma points, the post transformation mean is calculated using Eq. (11)

$$\widetilde{st}_{(k|k-1)} = \sum\nolimits_{j=0}^{2L} W_j^{(m)} \times st_{(j,K|k-1)} \tag{11}$$

4. Assuming process noise Q to be additive, the error covariance is calculated as shown in Eq. (12)

$$P_x = Q + \sum\nolimits_{j=0}^{2L} W_j^{(c)} \times \left[st_{j,k|k-1} - st_{(j,k|k-1)}\right] \times \left[st_{(j,k|k-1)} - st_{(k|k-1)}\right]^T \tag{12}$$

5. Transforming the observations
   a. The transformed sigma points are propagated through an observation function $h()$

$$y_{(k|k-1)} = h\left(st_{(k|k-1)}\right) \tag{13}$$

   b. With the transformed sigma points, the predicted output is calculated as shown in Eq. (14)

$$\tilde{y}_{(k|k-1)} = \sum\nolimits_{j=0}^{2L} w_j^m \times y_{(j,k,|k-1|)} \tag{14}$$

   c. With the transformed sigma points, the output covariance is calculated using Eq. (15)

$$\begin{aligned} p_y =& R + \sum_{j=0}^{2L} W_j^{(c)} \times \left[y_{(j,k|k-1)} - \tilde{y}_{(k|k-1)}\right] \\ & \times \left[y_{|j,k|k-1|} - \tilde{y}_{(k|k-1)}\right]^{\mathrm{T}} \end{aligned} \tag{15}$$

   d. The cross-covariance between state and output is obtained using Eq. (16)

$$\begin{aligned} P_{xy} =& \sum_{j=0}^{2L} W_j^{(c)} \times \left[st_{(j,k|k-1)} - \widetilde{st}_{(k|k-1)}\right] \\ & \times \left[y_{(j,k|k-1)} - \tilde{y}_{(k|k-1)}\right]^{\mathrm{T}} \end{aligned} \tag{16}$$

6. Measurement Update
   a. The Kalman gain K is calculated using Eq. (17)

$$K_k = p_{xy} \times p_y^{-1} \tag{17}$$

   b. With the Kalman gain K, the state matrix is updated using Eq. (18)

$$p_k = p_x + K_k \times p_y^{-1} \times k_k^T \tag{18}$$

   c. With the Kalman gain K, the covariance matrix is updated using Eq. (19)

$$\widetilde{st}_k = \widetilde{st}_{(k|k-1)} + K_k \times \left(y_k - \tilde{y}_{(k|k-1)}\right) \tag{19}$$

The values of parameters mentioned in the above UKF algorithm are as follows. The primary scaling factor α is set to unity. The tertiary scaling parameter κ is set to zero. The length L of the state is five. The process and measurement noise co-variances values are taken as $Q = 0.000000001$ and $R = 0.001$ respectively.

4 Results and Discussions

In this section the results of the simulations performed is given to analyze the performance of the algorithm. The simulation is performed for 100 realizations and 5 iterations. The values of scaling parameters are taken as $\alpha = 0$, $\beta = 2$, $\kappa = 0$. Figures 1, 2 and 3 represents the simulation results of the X and Y-coordinates of the position of WMR and the orientation of the WMR. The corresponding values of the desired and estimated position and orientation of the WMR are given in Tables 1, 2 and 3.

From Figs. 1, 2 and 3, it is evident that the algorithm tracks the desired path with high accuracy. Tables 1, 2 and 3 gives the values of the X-coordinate, Y-coordinate and the orientation angle .The maximum error in case of X-coordinate is 0.70780 and in case of Y-coordinate it is 0.51518 and with respect to it is 0.04035 which is very less compared to using EKF which is 1.5 [5]. The values show that with the proposed UKF algorithm, the robot can track the desired path with least error when compared to DR used alone. The localization of a mobile robot with the proposed system is highly reliable.

Table 1. X-coordinates of the position of the WMR

X (desired)	X (estimated)	Error
48.74875	48.97043	0.22168
−26.70180	−27.40960	−0.70780
26.32628	26.88819	0.56191
7.34365	7.56205	0.21570
−2.42144	−2.57112	−0.14968
−18.76140	−18.76319	−0.00179
32.99141	33.07856	0.08715

Table 2. Y-coordinates of the position of the WMR

Y (desired)	Y (estimated)	Error
32.64495	32.80648	0.16153
−27.65380	−27.13862	−0.51518
49.11877	49.39118	0.27241
−25.10505	−25.12583	−0.02078
−10.09247	−10.27290	−0.18043
−40.98339	−40.86732	−0.11607
11.98159	11.9863	0.00471

Table 3. Orientation of the WMR in radians

(desired)	(estimated)	Error
−1.54365	−1.54331	−0.00003
−1.48423	−1.47888	−0.00535
1.53279	1.53376	0.00097
−1.53279	−1.53438	−0.00159
−1.53348	−1.53631	−0.00283
−1.52684	−1.50249	−0.02435
1.50497	1.54532	0.04035

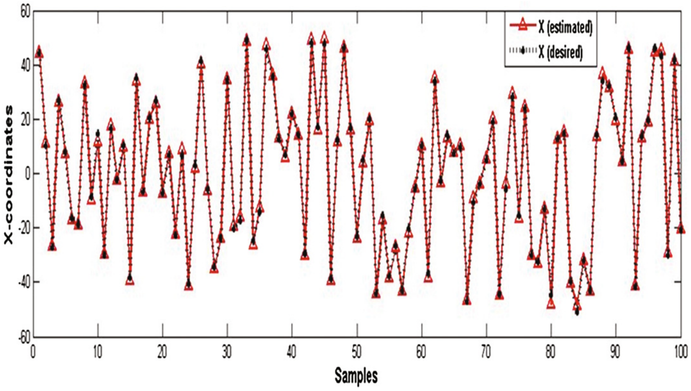


Fig. 1. Desired and estimated values of X-coordinates of position of the WMR

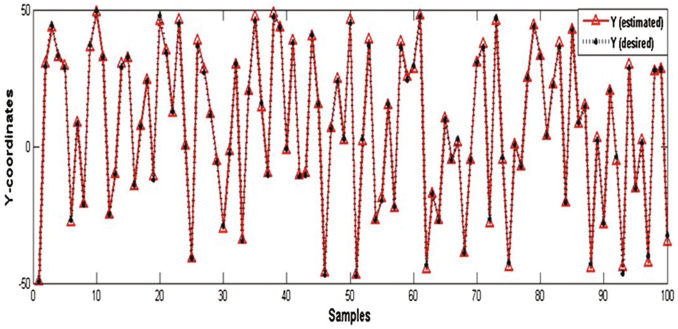


Fig. 2. Desired and estimated values of Y-coordinates of position of the WMR

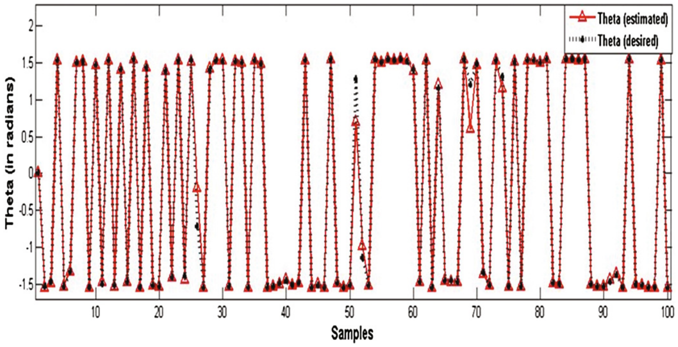


Fig. 3. Desired and estimated values of the orientation θ of the WMR

5 Conclusion

This paper has proposed a method to solve the problem of mobile robot localization using UKF. DR method of localization is having the effects of major errors due to slippage of wheels, tolerance rate of machine and surface roughness. The accumulation of errors due to these factors is highly reduced and making the localization highly reliable by using UKF along with DR method. From the simulation results it is observed that with the proposed algorithm, the robot can track the desired path with high accuracy. The proposed algorithm is robust and can be extended to other applications such as GPS tracking. Further improvement to the proposed system can be made my adaptively modifying the scaling parameter that affects the spread of sigma points and thus improves the overall filter tracking.

References

1. Amanatiadis, A.: A multisensor indoor localization system for biped robots operating in industrial environments. IEEE Trans. Ind. Electron. **63**(12), 7597–7606 (2016)
2. Chen, S.Y.: Kalman filter for robot vision: a survey. IEEE Trans. Ind. Electron. **59**(11), 4409–4420 (2012)
3. Chung, H.Y., Hou, C.C., Chen, Y.S.: Indoor intelligent mobile robot localization using fuzzy compensation and Kalman filter to fuse the data of gyroscope and magnetometer. IEEE Trans. Ind. Electron. **62**(10), 6436–6447 (2015)
4. Constanzi, R., Fanelli, F., Monni, N., Ridolfi, A., Allotta, B.: An attitude estimation algorithm for mobile robots under unknown magnetic disturbances. IEEE/ASME Trans. Mechatron. **21**(4), 1900–1911 (2016)
5. Faisal, M., Alsulaiman, M., Hedjar, R., Mathkour, H., Zuair, M., Altaheri, H.: ZakariahM, Bencherif MA and Mekhtiche MA.: Enhancement of mobile robot localization using extended Kalman filter. Adv. Mech. Eng. **8**(11), 1–11 (2016)
6. Kim, J.H., Seong, P.H.: Experiments on orientation recovery and steering of an autonomous mobile robot using encoded magnetic compass disc. IEEE Trans. Instrum. Measure **45**(1), 271–274 (1996)

7. Kim, S.J., Kim, B.K.: Dynamic ultrasonic hybrid localization system for indoor mobile robots. IEEE Trans. Ind. Electron. **60**(10), 4562–4573 (2013)
8. Nath, T.G., Sudheesh, P., Jayakumar, M.: Tracking inbound enemy missile for interception from target aircraft using extended Kalman filter. In: Proceeding of Communications in Computer and Information Science, vol. 625, pp. 269–279. Springer (2016)
9. Song, K.T., Suen, Y.H.: Design and implementation of a path tracking controller with the capacity of obstacle avoidance. In: Proceeding of Automatic Control conference, pp. 134–139 (1996)
10. Tsai, C.C.: A localization system of a mobile robot by fusing dead-reckoning and ultrasonic measurements. IEEE Trans. Instrum. Measure **47**(5), 1399–1404 (1998)
11. Tuna, G., Gulez, K., Gungor, V.C., Mumcu, T.V.: Evaluations of different simultaneous localization and mapping (SLAM) algorithms. In: Proceeding of IECON 2012-38th Annual Conference on IEEE Industrial Electronics Society, pp. 2693–2698 (2012)
12. Vikranth, S., Sudheesh, P., Jayakumar, M.: Nonlinear tracking of target submarine using Extended Kalman Filter (EKF). In: Proceeding of Communications in Computer and Information Science, vol. 625, pp. 258–268 Springer (2016)
13. Wang, S., Chen, L., Gu, D., Hu, H.: Cooperative localization of AUVs using moving horizon estimation. IEEE/CAA J. Automatica Sinica **1**(1), 68–76 (2014)

An Analysis on the Influence that the Position and Number of Control Points Have on MLS Registration of Medical Images

Hema P. Menon(✉)

Department of Computer Science and Engineering,
Amrita School of Engineering, Coimbatore, Amrita Vishwa Vidyapeetham,
Amrita University, Coimbatore, India
p_hema@cb.amrita.edu

Abstract. In this paper an analysis on the influence that the selection of fiducial points has on the Moving Least Square registration of medical images has been presented. MLS is a point based method which needs selection of fiducial (control) points. Here the mapping is weighted by the distance of current pixel from the selected point. Hence it is deemed significance to investigate on the effect that the position and number of the selected control points have on registered image. The analysis is done by manually selecting the points from rigid and non-rigid regions, near and far off regions from the two images and by also varying the number of points. To assess the results comparison has been done with the TPS registration by computing the TRE.

Keywords: Moving least squares (MLS) · Image registration · Medical images · Feature points · Target registration error (TRE)

1 Introduction

The process for image registration involves finding the best deformation field to align the two images under consideration. To register these images, features or fiducial points are selected from the images and correspondence is established between them. Once the correspondences are known, a mapping function is found to align source image to the geometry of the reference image. This can be done by either assuming that the transform is known or by assuming that the points are known and finding the other parameters. The second method is followed in this work.

The research in registration has gone a long way and many researchers have contributed a lot in this field. Images can be registered using either pixel intensity values [1] or based on the feature extracted from the images [2]. Both cases are applicable to rigid and non-rigid images. Intensity based methods are performed in spatial domain and works directly on pixel intensities [3, 4]. They are sensitive to intensity variations like background changes, noise, brightness, etc.

For medical images generally the feature based registration is used wherein features are extracted from each image and then the correspondence is matched using any measure like mutual information or entropy [5, 7]. Tian et al. [6] proposed a retinal

S.M. Thampi et al. (eds.), *Advances in Signal Processing and Intelligent Recognition Systems*, Advances in Intelligent Systems and Computing 678,
DOI 10.1007/978-3-319-67934-1_5

image registration framework, in which the features like vascular bifurcations are detected and registration is performed using Harris PIIFD. Zheng et al. [7] discusses registration of retinal images using a salient feature region computation method. Tsai et al. [8] uses an edge driven dual–bootstrap Iterative Closest Point Algorithm for registration. Modifications on the IPC algorithm have been proposed by many researchers for various applications [9, 10]. A validation of the general methods for image registration on different subjects obtained using the different modalities have been discussed by Wood [11]. Rigid registrations are generally used in applications dealing with bony data like CT image [12, 13]. A detailed theory on various aspects of medical image registration including algorithms, their validation and applications can be found in Fitzpatrick et al. [14]. Fitzpatrick [14] also discusses the ways of measuring the registration accuracy by predicting the registration error, by computing the Target Registration Error (TRE) and the Fiducial Registration Error (FRE). Registration is a very important pre-processing step for fusion of images [15, 16].

1.1 Registration Using MLS

The main challenge in registration of images is estimating the unknown parameters present in the problem like the transformation function, the mapping parameters and the correspondence points which are all unknown in this case. An illustration of the registration process is shown in Fig. 1. General approaches are to start by initializing any of them and then tuning the parameters accordingly. In this work the transformation function used is the Affine Transformation with the as-rigid-as-possible constraint. The moving least square error is computed after each of the iterations and the mapping parameters are tuned based on the optimization function. Here, the correspondence points are assumed to be known and then using them the mapping parameters are estimated. To assess the influence of the correspondence points, also known as control points, feature points, fiducial points or pivot points, on MLS registration, points are selected manually from both the target and the source images.

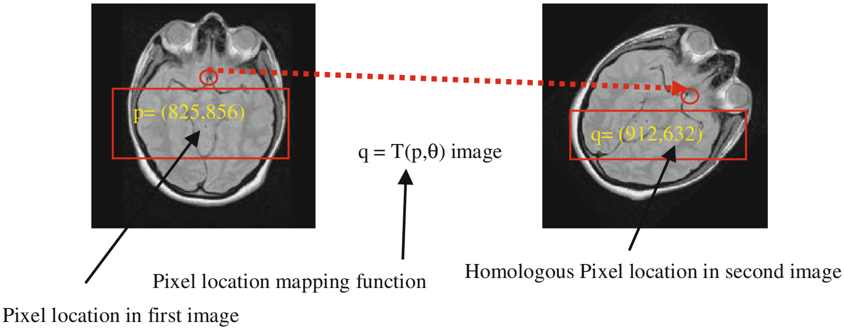

p= (825,856)
q= (912,632)
q = T(p,θ) image
Pixel location mapping function
Homologous Pixel location in second image
Pixel location in first image

Fig. 1. The registration process

2 Fiducial (Control) Point Selection and Its Influence on MLS

A set of points are manually selected from both the images. Figure 2 shows a sample of the point selection wherein 4 points were selcted from the source and target images. Let p = $\{x_{pi}, y_{pj}\}$ and q = $\{x_{qi}, y_{qj}\}$ be the coordinates of the selected set of points. The coordinate values are also given for the points. On observing the values it can be seen that the coordinate values for the same structural position in the source and target images are different. This difference has to be minimised to align both images and is the key task of regisrtration.

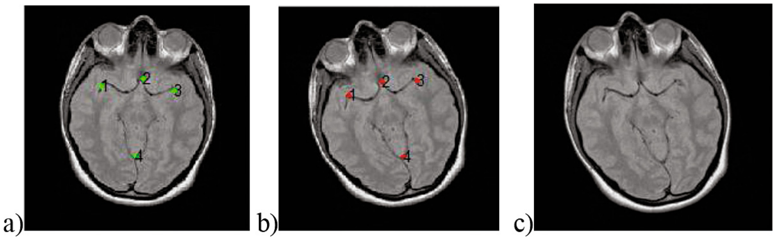


Fig. 2. (a) and (b) Manually selected fiducial points from source and target images respectively. (c) MLS registered source.

$$P = \{(87.4305, 92.2219)(135.3449,\ 84.0080)(170.9385,\ 97.6979)(127.1310, 171.6230)\}$$

$$q\ =\ \{(81.5213,\ 100.5851)\,(119.6489,\ 85.6064)\,(159.1383,\ 84.2447)(144.1596, 170.0319)\}$$

Let r = $\{x_{ri}, y_{rj}\}$ be the coordinates of the same structural points from registered source shown in Fig. 2(c). The values are as given below.

$$r\ =\ \{(82.4513,\ 99.5851)(118.5489,\ 84.6064)(157.7363,\ 86.6447)(142.1596,\ 168.9319)\}$$

On observing the values it can be seen that there is a small difference in the coordinates of the target and registered source. This is because while selecting the 4 points there might be a small error that is called as localization error, which has occurred since the points are selected manually. This error is called as the fiducial localization error (FLE). This in-turn effects the computation of TRE and hence normally the FLE is assumed to be nil during TRE calculation. To know if the image has been registered correctly the TRE is calculated and was found to be 0.9876 for the image in Fig. 2(c).

2.1 Results and Analysis of Influence of Number and Position of Control Points

Experiments were conducted on MRI and CT images by selecting different number of points and from varying locations. Points were selected from rigid and non-rigid regions in an image and also near and far off points. A large number of control points and also minimum number like 2, 3, 4, 6 etc. were selected. Such an analysis is very essential when considering medical images because

(i) Availability of many control points from both images under consideration may be difficult
(ii) Also the fact that the time complexity increases as the number of points taking part in the computation of the transformation function increases.

Hence it is important to see if accurate registration can be performed with as minimum control points as possible. The analysis has been conducted on MRI and CT images.

2.1.1 Analysis on Registration of MRI Images

The result obtained with MLS registration for some of the data sets with fewer control points is given in Table 1. To compare the efficiency of the MLS method, the same image pairs were given to TPS registration and the comparison results are given in Table 2. The analysis done by selecting minimal 2 points from different regions of images are shown. Points have been selected from rigid and non rigid regions from the image. On observing the TRE values in Table 2 it can be inferred that the registration error is less compared to TPS method and also that wherever the points be the average error is around 1.6.

Here it can be observed from Table 2 that the image registered using TPS is distorted, i.e., the registration is not perfect in some cases depending on the position of the selected points, especially in cases when 2 and 3 control points are selected. The distortion is less or not there along the axis in which the points lie.

On closely observing the TRE values for MLS it has been inferred that, TRE is lesser when more number of points are selected. In cases were 2 to 8 point are selected the error is dependent on the position of the selected points also. If all the selected points are nearby then the error is more compared to the same number of points being selected such that are well spaced from each other. Also if the points all lie on the same side or quadrant, then also the error is comparatively more. So the best way to select the points is such that they cover points from all quadrants and are slightly far apart. If the points are selected in this fashion then the TRE will be between 0.9778 and 1.275 for MLS registration.

To assess this experiments were conducted on CT images also as its characteristics differ from that of MRI images. The results with respect to this are discussed in Sect. 2.1.2. CT has got more rigid structures than MRI images which show more of tissue level details and can have local deformation in them.

From the Table 2 it can be seen that with only 1 point selected MLS does not give ant registration result. But with a minimum of 2 points registration is possible.

Table 1. MRI image pair showing the control points selected and the MLS transformed source image.

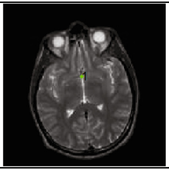
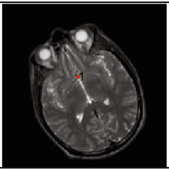
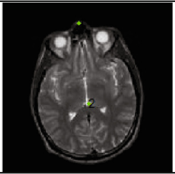
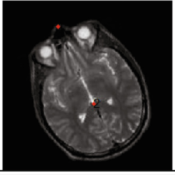
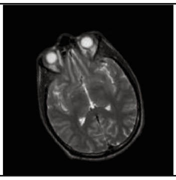
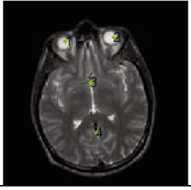
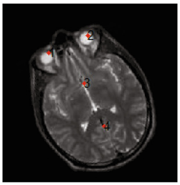
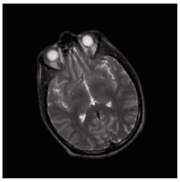
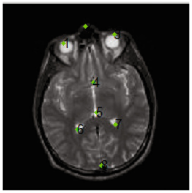
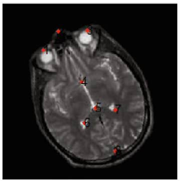
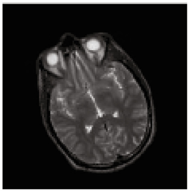
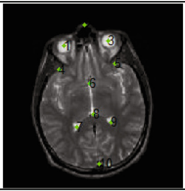
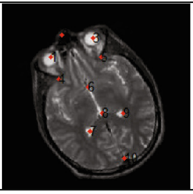
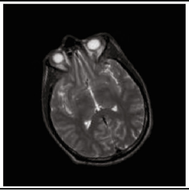
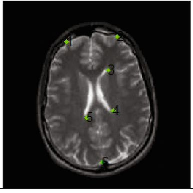
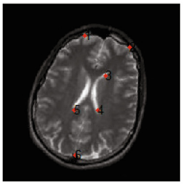
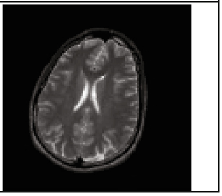
No. of points selected	Source Image	Target Image	Registered Source
1			
2			
4			
8			
10			
6			

Table 2. Analysis for 2, 3 and 6 points selected from different regions on MRI images.

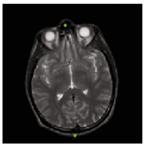
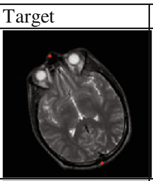
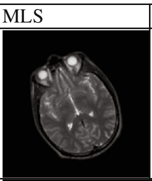
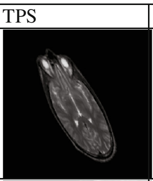
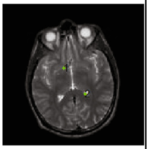
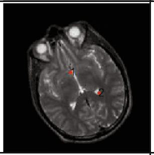
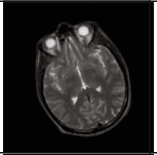
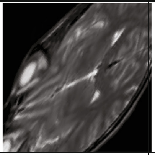
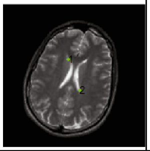
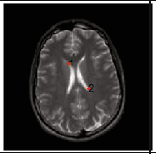
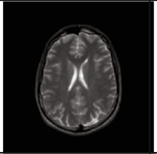
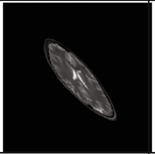
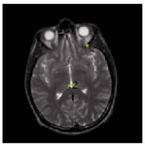
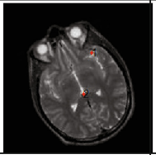
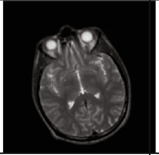
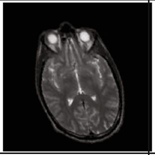
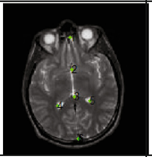
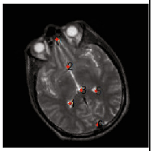
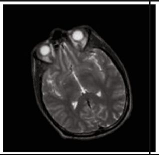
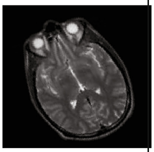
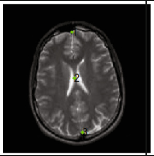
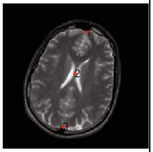
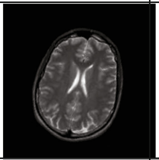
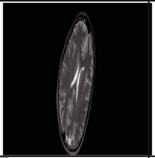
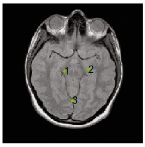
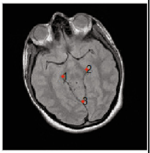
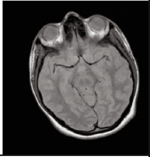
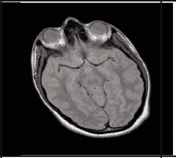
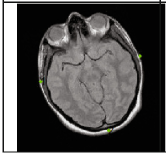
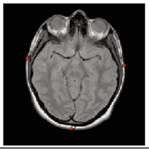
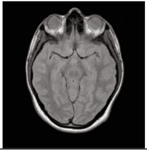
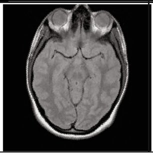
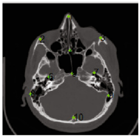
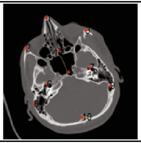
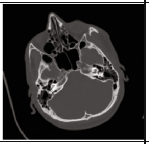
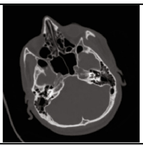
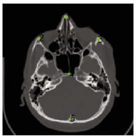
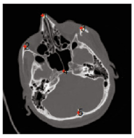
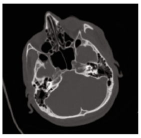
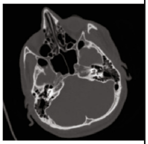
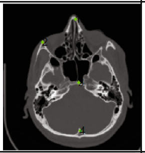
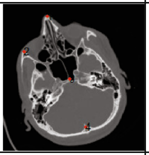
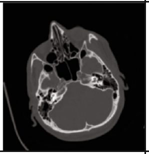
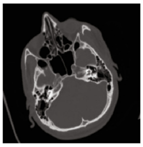
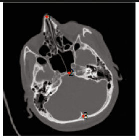
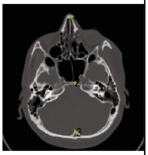
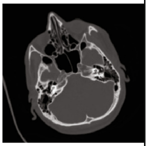
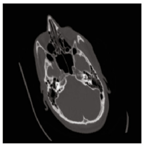
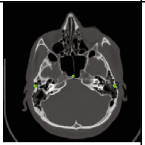
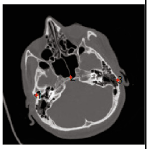
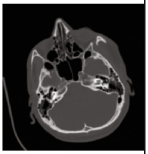
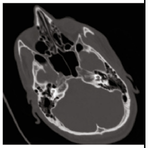
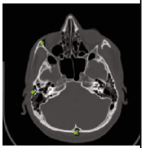
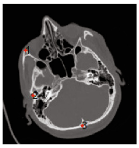
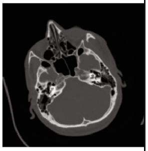
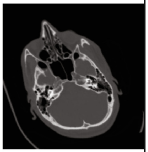
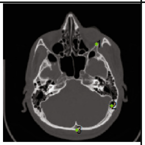
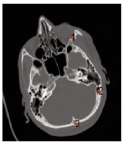
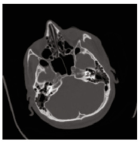
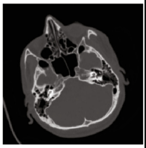
Source	Target	MLS	TPS	TRE_{MLS}	TRE_{TPS}
				1.6021	2.4672
				1.6306	3.9892
				1.6398	2.8656
				1.5798	2.2063
				1.4367	1.7761
				1.3231	2.3563
				1.4876	2.1065
				1.3973	2.2876

Table 3. Analysis for varied number and position of selected points on CT images.

Source Image	Target Image	MLS	TPS	TRE_{MLS}	TRE_{TPS}
				1.2033	1.3760
				1.3998	1.7982
				1.3002	1.8001
 Points from center				1.5869	2.5356
 Points from middle				1.6540	2.4098
 Points from Left Half				1.5998	2.2702
 Points from Right Half				1.5765	2.2675

2.1.2 Analysis on Registration of CT Images

To analyze the efficiency of the MLS registration experiments were also conducted on CT images also. A different number of control points were selected from different regions and the results were compared with TPS. A sample of the results for 10, 5, 4 and 3 points are shown in Table 3.

2.1.3 Inferences

Based on similar experiments conducted by selecting various number of control points from different regions of the image the following inference is obtained.

1. From the analysis performed it is inferred that the MLS registration performs well in case of fewer points also.
2. However the minimum number of points required is atleast 2.
3. When only 1 point is selected there is no registration output.
4. The directions along which the points are selected are registered more accurately than the other regions.
5. Hence, while selecting points, they must be selected such that the structure of the image is roughly maintained.
6. The points in case of less number of points being selected, the selection must be from far off regions, as MLS calculates weights based on the distance of the current point from the control point selected for a lower TRE.
7. Similar results were got in case of both MRI and CT images.
8. In cases where it is possible to get only a few points from the images under consideration, it is hence suggested that MLS is a better approach for registering both Rigid and Non-Rigid images.

2.2 Quantitative Analysis

To analyze the influence that the fiducial points have on the performance of the MLS registration the metric used was the Target Registration Error (TRE). The TRE obtained was compared with the TPS method. In order to compute the variations in the TRE we calculated the standard deviations under both the methods as shown in

Table 4. Comparison of TRE for MLS and TPS registration

No. of control points selected	Standard deviation of target registration error	
	MLS registration	TPS registration
3	1.6432	2.2689
4	1.3533	1.6567
5	1.4065	1.7250
6	1.4032	1.7156
8	1.2002	1.4132
10	1.2356	1.3786
21	1.1056	1.2442
40	1.0090	1.1104

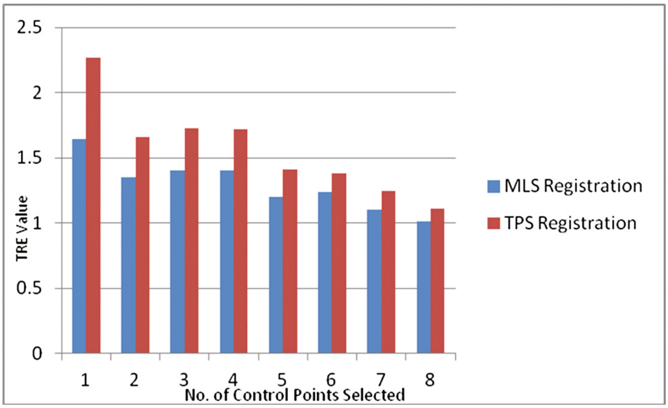


Fig. 3. Graphical representation of the TRE versus no. of points selected

Table 4. Since the ideal value of TRE is generally accepted as 1, this method of registration appears to be superior to that of the TPS TRE, because of the lesser standard deviations observed in the MLS TRE. This is more prominent when the number of control points taken is lesser. Form the graphical plot given in Fig. 3 it the effect that the number of control points has on the MLS and TPS registration can be clearly seen. As the number increases the TRE decreases and the performance of both the registration methods are almost similar. But for lesser number of points the MLS outperform the TPS registration in seen in case of MRI and CT brain data.

3 Conclusion

In this work a thorough analysis on the influence that the position and the number of fiducial points selected from the reference and floating images have on the MLS registration has been performed. Such an analysis is deem of significance as in case of medical images availability of large number of points for correspondence matching may not be possible due to the nature of the images available. In this work experiments were conducted on MRI and CT images and the results obtained using MLS registration was compared with TPS to assess the effectiveness of the MLS algorithm. This was done by computing the TRE for both the registered sources. It has been inferred from this analysis that MLS can perform better even when a minimum of 2 or 3 correspondence points are available from the images. However the factor to be taken care of in such case is that the two points are distantly placed. When more number of points are available the position of the points does not matter for MLS registration.

References

1. Pluim, J.P.W., Maintz, J.B.A., Viergever, M.A.: Mutual information-based registration of medical images: a survey. IEEE Trans. Med. Imaging **22**, 986–1004 (2003)
2. Butz, T., Thiran, J.P.: Affine registration with feature space mutual information. In: Medical Image Computing and Computer-Assisted Intervention, vol. 2208, pp. 549–556. Springer-Verlag (2001)
3. Maes, F., Collignon, A., Vandermeulen, D., Marchal, G., Suetens, P.: Multimodality image registration by maximization of mutual information. IEEE Trans. Med. Imaging **16**, 187–198 (1997)
4. Rodríguez-Carranza, C.E., Loew, M.H.: A weighted and deterministic entropy measure for image registration using mutual information. Med. Imaging. Image Process. **3338**, 155–166 (1998)
5. Ioannides, A.A., Liu, L.C., Kwapien, J., Drozdz, S., Streit, M.: Coupling of regional activations in a human brain during an object and face affect recognition task. Hum. Brain Mapp. **11**(2), 77–92 (2000)
6. Tian, J., Lee, N., Theodore, R., Smith, A., Laine, L.: A partial intensity invariant feature descriptor for multimodal retinal image registration. IEEE Trans. Biomed. Eng. **57**(5), 1707–1718 (2010)
7. Zheng, J., Tian, J., Deng, K., Dai, X., Min, X.U.: Salient feature region: A new method for retinal image registration. IEEE Trans. Inf. Technol. Biomed. **15**(2), 221–232 (2011)
8. Tsai, C., Li, C., Yang, G., Lin, : The edge-driven dual-bootstrap iterative closest point algorithm for registration of multimodal fluorescein angiogram sequence. IEEE Trans. Med. Imaging **29**(3), 636–649 (2010)
9. Almhdie, A., L´eger, C., Deriche, M., Lédée, R.: 3D registration using a new implementation of the ICP algorithm based on a comprehensive lookup matrix: application to medical imaging. Pattern Recogn. Lett. **28**(12), 1523–1533 (2007). Elsevier
10. Pan, M-s, Tang, J-t, Rong, Q-s, Zhang, F.: Medical Image registration using modified iterative closest points. Int. J. Numer. Method Biomed. Eng. **27**(8), 1150–1166 (2011)
11. Woods, R.P., Grafton, S.T., Holmes, C.J., Cherry, S.R., Mazziotta, J.C.: Automated image registration: I. General methods and intrasubject, intramodality validation. J. Comput. Assist. Tomogr. **22**(1), 139–152 (1998)
12. Andreetto, M., Cortelazzo, G.M., Lucchese, L.: Frequency domain registration of computer tomography data. In: proceedings of 2nd International Symposium on 3D Data Processing, Visualization and Transmission (3DPVT 2004), pp. 550–557 (2004)
13. Heger, S., Portheine, F., Ohnsorge, J.A.K., Schkommodau, E., Radermacher, K.: User interactive registration of bone with A-mode ultrasound. IEEE Eng. Med. Biol. Mag. **24**(2), 85–89 (2005)
14. Fitzpatrick, J.M., West, J.B., Maurer, C.R.: Predicting error in rigid-body point based registration. IEEE Trans. Med. Imaging **17**(5), 694–702 (1998)
15. Moushmi, S., Sowmya, V., Soman, K.P., Empirical wavelet transform for multifocus image fusion. In: Advances in Intelligent Systems and Computing. vol. 397, pp. 257–263 (2016)
16. Sruthy, S., Latha, P., Sasi, A.P.: Image fusion technique using DT-CWT. In Proceedings—2013 IEEE International Multi Conference on Automation, Computing, Control, Communication and Compressed Sensing, iMac4s 2013, Kerala, pp. 160–164 (2013)

Component Characterization of Western and Indian Classical Music

Shivam Sharma(✉), Seema Ghisingh, and Vinay Kumar Mittal

Indian Institute of Information Technology Chittoor, Sri City, AP, India
{shivam.sharma,seema.ghisingh}@iiits.in, DrVinayKrMittal@gmail.com

Abstract. Regular pitch detection algorithms are known to be immensely useful for speech source analysis. Their utility is not as reliable when processing polyphonic acoustic mixtures like Music. This is an investigative study of music components like rhythm, accompaniment and Lyrical-voicing, that is seen as a critical task towards targeted music component identification and processing. Popular music forms like Western and Hindustani Classical are considered for our study dataset. For Western cases, comparative preliminary analysis of the spectral characteristics like Harmonics and Energy is done towards characterization of Music region against that of Lyrics-music mixture. F_0 contour analysis for these regions, using Autocorrelation and Zero frequency filtering indicates the utility of the latter in Lyrical-voicing onset identification. Short-time spectral analysis leads to the distinctive understanding about the Harmonic structure according to the music polyphony. Strength of Excitation is found to be insightful towards characterizing sounds like base sounds, prominent in percussion instruments. For study on Classical music, F_0 contour analysis using raw signal and LP Residual elucidate the characteristic average pitch effect, which comes out to be higher for the Alaap region in case of Female artists and Lyrics composition regions for the Male artists, giving cues towards the applications like Raaga identification and summarization. The analysis of the excitation source features for various music components done in this work present some insightful observations and clues towards effective Music component processing.

Keywords: Western · Classical · Pitch · Harmonics · Energy · Raaga

1 Introduction

Music as known to everyone has its roots from the Classical styles from various cultures. In western form the music knowledge is known to have been documented from early 500 ADs whereas, the earliest reference to the Hindu classical music theory is known to trace back to 400 BC. A comprehensive understanding of today's music is achieved by studying both types. There are various prospects of studying the singing vocalisation, for instance component enhancement, separation, melody tracking, etc. Considering this as our central theme, we have

S.M. Thampi et al. (eds.), *Advances in Signal Processing and Intelligent Recognition Systems*, Advances in Intelligent Systems and Computing 678,
DOI 10.1007/978-3-319-67934-1_6

attempted to characterize the music components in both time and frequency domains, for both of these styles.

The frame level feature vectors were processed using Gaussian Mixture Model focussing on Singing Voice in [13]. The aperiodicity in 'Noh' Voices was studied in [7] using *modZFF*. Also, along-with the consideration of analysis by synthesis using two synthetic AM/FM sequences for excitation, and Saliency as a measure of pitch perception, this technique has resulted in encouraging results when worked upon highly varying source excitation characteristics for sounds like verbal/non-verbal paralinguistic sounds as in [6]. Some fundamental work in blind music-source separation using repetition was done in [12]. Modelling of Non-vocal accompaniments was achieved after the component segmentation in [10]. Another Voice/Non-voice segmentation was done using Hidden Markov Model and used as a basis along with an inference method to derive vocals from a mixture in [4]. The significance of the repetition for the positive effect of music on human psyche was highlighted in [8].

Variations in singing voice were studied using conventional pitch detection algorithms (PDAs) in [14]. Autocorrelation was used to extract the F_0 contours. Pre-emphasis was found useful in case of sounds having high frequency. Since, the approximation quality was limited for the LP (Linear Prediction) Residual based contour extraction, other techniques that could capture the source excitation information in a better way were required to be evaluated. The motivation for the Singing characterization came in while performing experiments related to Music Source Separation using Autocorrelation based *static* background modelling techniques in [15]. It was observed that even with the dynamic music modelling techniques, targeted components processing is highly required.

Autocorrelation and Zero Frequency Filtering (ZFF) were additionally evaluated on a music dataset of Western form by Female artists, as part of the current work for the purpose of F_0 contour extraction. The Harmonic structures of different accompaniments along-with the voicing parts are studied in detail. Also assessed are the Harmonic structure of the mono-pitched and multi-pitched sound sources, providing cues towards better source excitation information extraction and further understanding about the targeted source and objective.

As for the Classical Music, the voicing in the Singing constituent of a music mixture is studied for the Pitch behaviour during *Lyrical composition* and *Alaap*. This is achieved using Autocorrelation of Raw signal and LP Residual followed by ZFF technique. These were also characterized using the Energy values. The results obtained are encouraging towards singing labelling and better understanding of the Raaga structure (onsets).

This paper is organized as follows. Music dataset is discussed in Sect. 2. In Sect. 3, Signal processing methods and features explored are described. Experimental details are included in Sect. 4. Acoustic analysis of Western and Classical music is discussed in detail in Sect. 5. Finally the paper is concluded and summarised in the last section.

2 Music Dataset

One of the forms of music considered for this work is Western music form. It comprises of various genres like Pop, R&B and Rock, etc., that have their roots in the cultural evolution from different parts of the western world. For instance from slow and groovy beats in R&B's to raw and pacy Country music. The naming convention is set in a manner as shown in Fig. 1.

For the study on classical music, the data was obtained from an online resource,[1] from which 5 male and 5 female music files were studied. The time duration of each raga file is approximately 6 s. Raaga files have two parts - *alaap* and *composition*, for both male and female speakers. Therefore, a total of 100 raga files were used for the purpose of characterization in this study. A brief description of some keywords is given as below:

- *Alaap:* It means a dialogue or conversation. Alaap is a dialogue between the musician and theh Raag. Alaap reflects the depth, the temperament, creativity and training of the musician.
- *Raaga:* Modes which express different moods in certain characteristic progressions, with more emphasis placed on some notes than others.

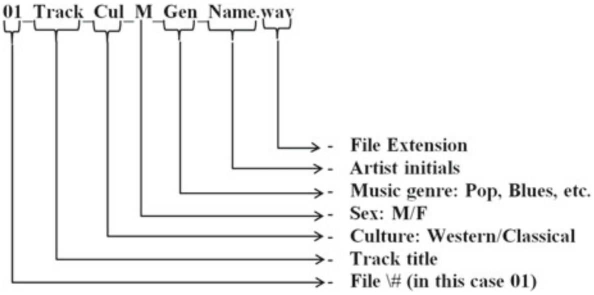

01_Track_Cul_M_Gen_Name.wav
- File Extension
- Artist initials
- Music genre: Pop, Blues, etc.
- Sex: M/F
- Culture: Western/Classical
- Track title
- File \# (in this case 01)

Fig. 1. Naming Convention for the Music Files.

3 Methods and Features

The Signal Processing Methods used, along with the Features explored are enlisted below,

(a) *Short-time Fourier Analysis*: Short Time Fourier Transform (STFT) is used to determine the sinusoidal frequency and phase contents of the local segments of the music mixture signal by processing it in the frequency domain.

[1] Ocean of Ragas: A dedicated collection of 1800+ Ragas of Hindustani Classical Music [Online]. http://www.oceanofragas.com.

This can be expressed as,

$$X(\tau, \omega) = \sum_{n=-\infty}^{n=+\infty} x[n]w[n-m]e^{-jwn} \tag{1}$$

Where, x[n] is the signal and w[n] is the window function [9]. Harmonics structure and Time-Frequency information are obtained using this.

(b) *Autocorrelation*: Autocorrelation provides a measure of the similarity between the waveforms of the time functions [3]. For the music signal x(n), whose correlation function is defined as

$$\begin{aligned} r_x(m) &= E[x(n)x(n+m)] \\ &= \lim_{N\to\infty} \frac{1}{2N+1} \sum_{n=-N}^{n=N} x(n)x(n+m) \end{aligned} \tag{2}$$

Pitch information about the sound sources are extracted using this technique.

(c) *Linear prediction (LP) analysis*: The prediction error *e(n)* or *LP Residual* can be computed by the difference between the actual sample *s(n)* and the *predicted* samples $\hat{s}(n)$ [11], and is given by $e(n) = s(n) - \hat{s}(n)$, i.e., $e(n) = s(n) + \sum_{k=1}^{p} a_k s(n-k)$. LP Residual is used as an alternative excitation source for capturing the Glottal activity, primarily for validation purpose.

(d) *Zero-Frequency filtering*: The essence of Zero-frequency filtering lies in computing the output of the cascade of two zero-frequency resonators [18], which is equivalent to four times successive integration of $x[n]$,

$$y_1[n] = -\sum_{k=1}^{2} a_k y_1[n-k] + x[n], \tag{3}$$

Where, $a_1 = -2, and\ a_2 = 1$. Pitch information and intensity of Glottal closure are derived using this technique.

(e) The vocalization in case of concerned regions in the Classical Music dataset are analysed using Signal Energy.

From now on for conciseness, Autocorrelation function, Linear Prediction and Zero-frequency filtering may be used interchangeably with ACF, LP and ZFF respectively.

4 Experimental Details

A 4096 point FFT based spectrogram was used for preliminary acoustic analysis, in the popular speech analysis tool called Wavesurfer [16]. The input music signals were observed post application of the voice activity detection (VAD) based upon first-order Markov modelling of the speech content [2,5,17] (marked in red, Fig. 5). A popular SOA algorithm YIN [1] was used as pitch references here, shown in different colours at the background of contour sub-plots. Short-time

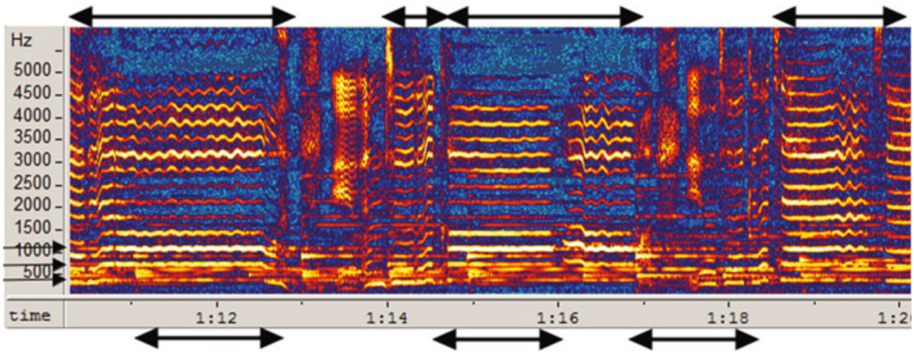


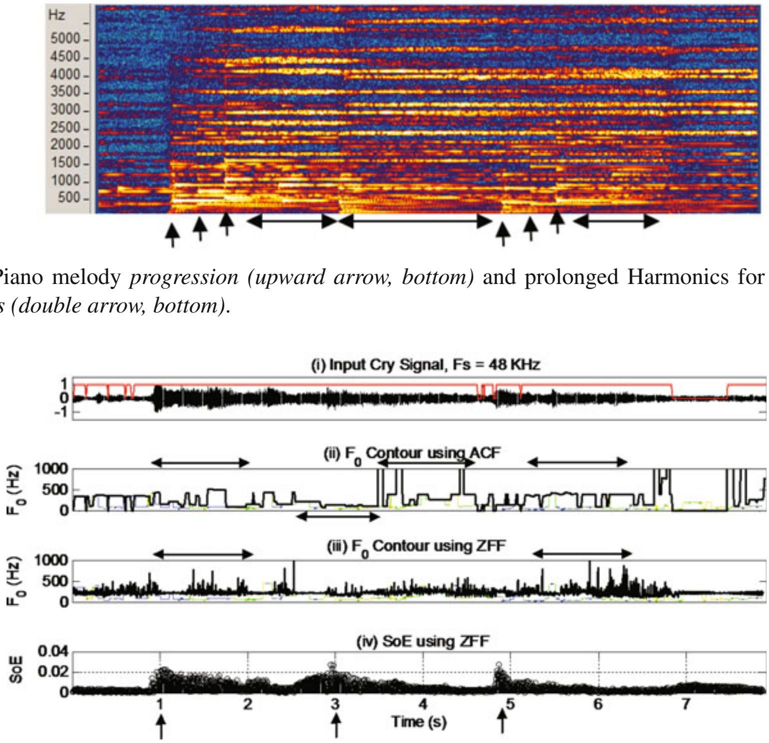
Fig. 2. Spectrogram of a typical western song. Please observe the changes in the arrow marked regions: Harmonic variations for *Lyrical-voicing (top)*, Flat Harmonics of *Piano Notes (bottom)* and Continuous *Piano Harmony (Left-edge)*.

analysis happens for both Autocorrelation and LP Analysis. Configuration: 30 ms and 10 ms Segment Size and Shift for F_0 extraction steps, 12 as LP order for LP Analysis. The experiments showed varied results for the female singers, especially with the ACF and LP residual based methods, hence ZFF was considered. The pitch variation for ZFF estimates was analysed using Standard-deviation. Harmonic Analysis was done using an app *Spectrum Analyzer*, primarily for observing the polyphony characteristics. Configuration: 44.1 KHz Sampling rate, 8192 FFT size.

5 Acoustic Analysis of Western and Classical Music

5.1 Characterization of Western Music

The preliminary spectral analysis for a western pop song called "My Love" by a Female artist *Sia* was done using a spectrogram and is shown in Fig. 2. Being a slow paced and low complexity accompaniment case, the spectral effects and the artefacts could very well be studied. The monotonous harmonics interleaved in between other components, prominently for the first half of the Frequency scale shown in Fig. 2 (please observe arrow marked regions at the bottom), represent the flat spectral characteristics of Piano Notes being played at onsets $71\,s, 73\,s, 74.7\,s \;\, and \;\, 77\,s$, as part of the lead accompaniment melody. Other accompaniment which is a keyboard harmony provided at the background, is observed to be continuous throughout (please observe right pointing arrows on the left edge). The lyrical voicing regions can be clearly distinguished on the basis of harmonics that are not as uniform and clean as for the accompaniments. Another key aspect of the voicing called as *vibrato*, can be observed as oscillating and fluctuating harmonics (please observe arrow marked regions at the top) in Fig. 2. The significant variation induced by the voicing improvisations manifest as prominent variations in the source excitation characteristics.



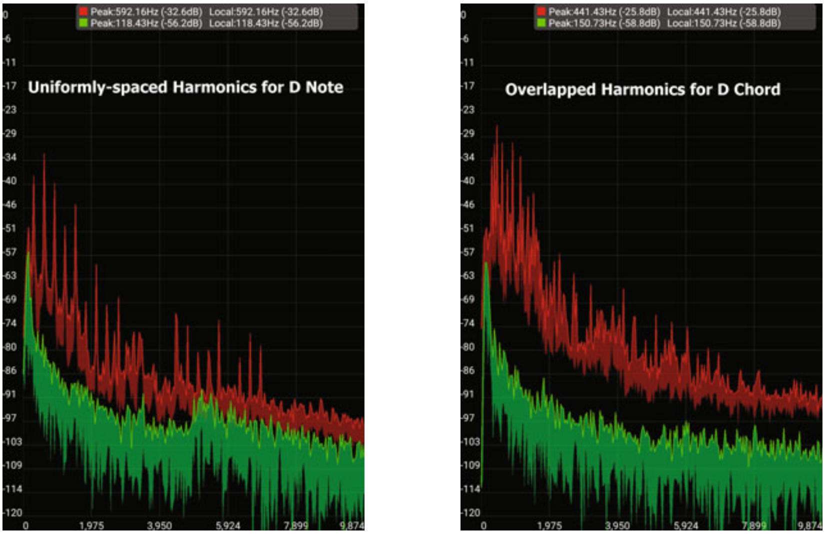
(a) Piano melody *progression (upward arrow, bottom)* and prolonged Harmonics for *singular notes (double arrow, bottom).*

(b) Changes in excitation source marked with arrows: (ii) and (iii) Lyrical-voicing mix (top) and (iv) SoE at Base effect onset (bottom).

Fig. 3. Spectral Analysis, F_0 contour and SoE comparison for Music excerpt.

Acoustic Analysis Using F_0 Contour

A. *Music Characterization:* The transitions of the harmonics in the lower half of the visible frequency range are due to the Piano melodic piece, whereas the relatively monotonous harmonic patterns visible in the higher Frequency regions (Fig. 3a), are because of the high pitched violin. This monotonous behaviour is also reflected in the F_0 contour, for the regions having only the violin as accompaniment, as can be seen at around 110 Hz in Fig. 3b, whereas for the melody onsets that are accompanied by a blend of base and mid-scale chords, along with the violin, F_0 estimation becomes random. These are the regions that will require further heuristics involving the harmonics structure, sub-band spectral energy, etc. Also, there appears to be octave errors by the YIN estimates near 3.25 s for instance, generating a contour at around 55 Hz, the corresponding signatures of which are not found in either of the analysis tools being used here. ACF gives better estimate of the dominant melody in a music mixture, whereas ZFF gives more insight during increased variations of lyrical part, which in the current form cannot be relied upon for a good melody approximation.

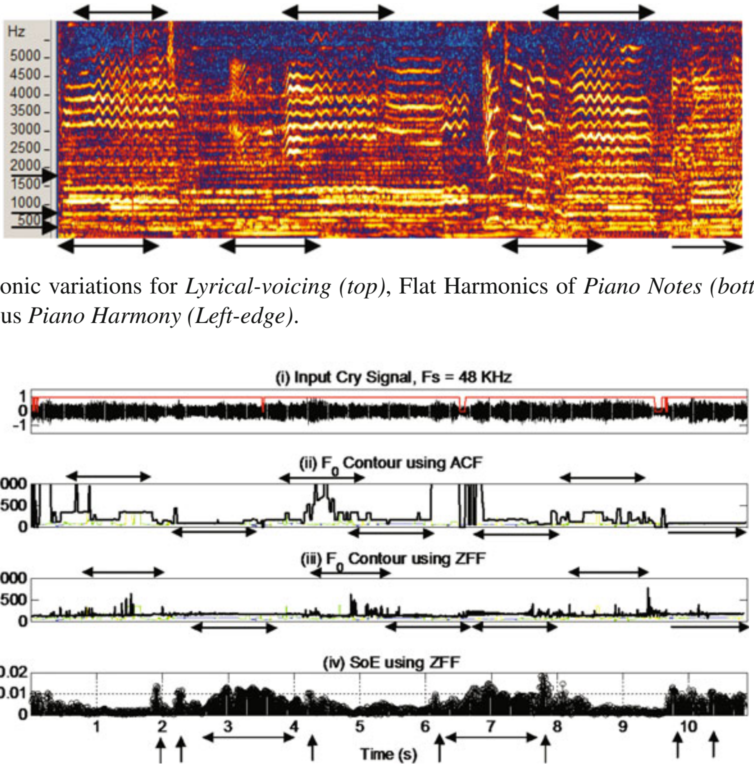


(a) Mono-pitched audio spectrum *(red, top)*. (b) Multi-pitched audio spectrum *(red, top)*.

Fig. 4. Illustration of the difference in Harmonic structure for (a) the mono-pitched sound with *uniformly spaced* Harmonics and multi-pitched sound mixture with *overlapping* Harmonics.

A comparison of Harmonic structure for Mono-pitched and Multi-pitched sound mixture is done. These were the audio input from a *Casio SA-35 Keyboard* with *4-Note maximum polyphony*. The *mid D note* was struck for generating a mono-pitched sound mixture whereas, *mid D-Chord* (triad) was used for creating a multi-pitched sound source. The magnitude spectrum was captured for these two test sound types and as can be observed from the Fig. 4 (please observe the patterns shown at the top, in Red), and the difference in Harmonic structure is clearly visible. For the mono-pitched sound, *uniformly spaced* Harmonics are obtained, whereas for the multi-pitched sound mixture, *overlapping* Harmonics are observed, which are not discernible right away. Probabilistic approaches like likelihood estimation can be utilised for deriving the sound source information in such scenario.

B. *Lyric-Music Mixture characterization:* The regions with violin and similarly pitched chorus generate correct F_0 contour approximates whereas, erroneous F_0 contour are generated at $5-6\,s$ because of the high tempo melody progression for Piano Notes resulting in greater overlap of the acoustic components (Fig. 5b). The contour estimates are random for Lyrical content. When Spectrogram is observed for the Lyrical content (Fig. 5a), the spectral characteristics give intense cues towards component identification in terms of Harmonic Behaviour. The Lyrics specific characteristics are clearly visible as highly fluctuating harmonics in the Higher Frequency ranges i.e., $> 3000\,Hz$. The harmonics at the key lyrics part here provides cues towards melody tracking, that Wavesurfer and YIN fails to capture. Also, for the same events when evaluated using a time domain based pitch detection algorithm (PDAs) like Autocorrelation and ZFF, in their current form cannot be relied upon for obtaining accurate pitch estimates. Although, ZFF based F_0 estimates do give an idea about the onset/offset of the regions with music and Lyric-music regions, based upon the *deviations* observed in the F_0 estimates. It can be clearly observed from Fig. 5b (subplot (iii)), Specific Onsets for the Lyrical-Voicing part with the Music is being identified, without the inclusion of the effects due to high-pitched vocalizations in the locality if any, as in case with the ACF based contour in Fig. 5b, and thus giving better insight towards the required onset information.



(a) Harmonic variations for *Lyrical-voicing (top)*, Flat Harmonics of *Piano Notes (bottom)* and Continuous *Piano Harmony (Left-edge)*.

(b) Changes in excitation source characteristics marked with arrows: (ii) and (iii) Lyrical-voicing mix (top) and (iv) SoE at Base effect onset and chorus vocals (upward and double sided arrow respectively, at bottom).

Fig. 5. Spectral Analysis, F_0 contour and SoE comparison for Lyrics-Music mixture excerpt.

Acoustic Analysis Using *SoE*. With the normal speech production mechanism, the idea is straight forward and it is the intensity of glottal closures, that we talk about when analysing using SoE, but when it comes to music component analysis, this technique has shown to be insightful towards highlighting the onset/offsets of the rhythm characterizing music components like music beat, base cover, and intense vocal renditions.

As can be seen from the SoE subplots for the Music excerpt in Fig. 3b, the entire visible piece has melodic transitions involving the Piano notes as the leading sound source visible at $1,\ 3\ and\ 5\,s$, as distinct surges in SoE values. These are the instants where the chords combined with the Base chords are struck. On a similar note, in Fig. 5b, the distinct SoE spike observed at the time instants $2,\ 4.2,\ 6.2,\ 8\ and\ 10\,s$, signify the base violin chord support whereas, the rest of the high surge points around $3.2\ and\ 7\,s$, having a bunch of high SoE points together in contrast with the *instantaneous surge* in the former case, are attributed either to the background chorus singing, or high pitched improvisations by the lead singer. Other experiments have also shown encouraging results for SoE based characterisation in case of percussive accompaniments, giving cues towards drum based music rhythm identification.

Observations in General

A. *Lyric-Music Mixture*: For many cases, resonant frequencies are clearly visible during the lyrical portion, helpful towards Lyrical region identification.

B. *Music Mixture*: SoE plays vital role in screening out some key music activities. It showed insightful results for the instruments like Piano, Drums, Guitar and even non music sounds like Clap and muti-pitch scenarios. The F_0 appears to be relatively more varying for Lyric-Music mixture as compared to just the music parts, hence giving more Standard deviation values as can also be observed from Fig. 6.

5.2 Characterization of Classical Singing

The techniques used here for the extraction of F_0 contour are Autocorrelation of LP Residual and ZFF, both giving insights towards voicing analysis in Classical music form. The region of interest *(ROIs)* in these types are alaap regions also known as improvisation part and lyrics part which basically contains more singing with words. On comparing F_0 extraction for both techniques, especially for female singers, the alaap regions have higher F_0 values as compared to the lyrics composition part. The difference found was in the range of $30 - 40\ Hz$ as seen in Table 1. However, the observations have different trend for male artists. Here, the lyrics parts have higher pitch values than alaap part. The differences in the pitch ranges are observed for the ROIs of the same track and artists. The F_0 differs for both the male and female artists for different parts of the classical music.

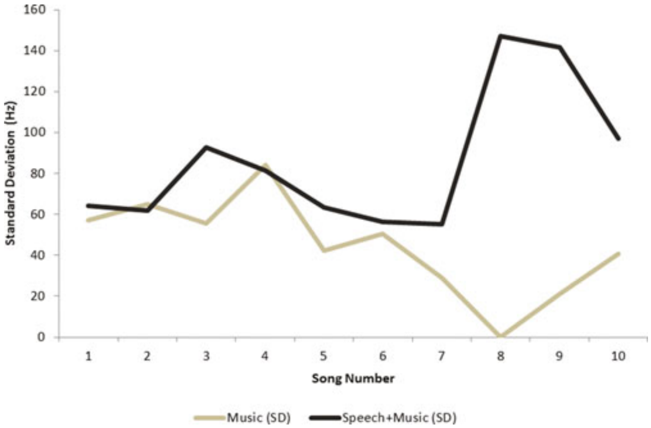


Fig. 6. Standard Deviation for the ZFF based Instantaneous F_0, compared for the Lyrics-Music Mix and just the Music excerpts.

Table 1. Mean F_0 values of 5 Male and 5 Female speakers for Indian raga: (a): Speakers, (b) and (c): F_0 using Autocorrelation of LP Residual for alaap and lyric composition (d) and (e): F_0 using ZFF for alaap and lyric composition.

	Autocorrelation		*AC of LP residual*	
(a) Speakers	*(b) alaap*	*(c) Lcomp*	*(d) alaap*	*(e) Lcomp*
S1 (M)	178.796	203.771	149.189	216.836
S2 (M)	167.414	210.606	165.520	223.774
S3 (M)	198.992	197.819	161.539	164.611
S4 (M)	205.876	234.428	176.895	234.180
S5 (M)	146.835	1715.532	175.852	191.790
S1 (F)	316.028	291.472	256.017	267.687
S2 (F)	289.693	260.387	251.865	249.163
S3 (F)	293.214	289.061	255.924	278.462
S4 (F)	333.121	324.441	263.824	333.284
S5 (F)	338.967	291.599	281.826	286.984

The energy gives the voicing information about different ROIs. For Female artists, the energy is higher for lyrics composition part as compared to that of alaap regions. Hence, it can be stated that during an alaap region there are instants where the voicing is not as traceable and becomes weak on reaching *high* notes. However, in case of lyrics the energy is found to be significantly distributed, hence giving higher values. On the other hand, in case of Male artists, higher modulation is introduced in the alaap region making it more tractable, as a result of which energy is higher in alaap regions too. The average Energy values for the Alaap and Lyrical composition regions for Male and Female artists respectively, are 0.071 *and* 0.076 for Male whereas, 0.100 *and* 0.109 for Female.

The key observations from the Experiments performed for the current work are included in Tables 2 and 3.

6 Summary and Conclusion

The acoustic analysis of the Music mixtures, using Spectrogram brought forth the significance of Harmonic patterns for finding excitation source information, as against the constrained utility of time domain pitch detection algorithms

Table 2. Analysis of Singing Voice in Western Music

S. #	*Key resulting observation*
1	Better approximate of the pitch contour in terms of the lead melody is given by the *Autocorrelation* method
2	F_0 extraction using ZFF is not as reliable for melody tracking, but it's deviation does gives cues for the presence of Lyrics-music mix as against just the Music as shown in Fig. 6
3	Lyrical voicing is observed to manifest in the form of highly varying and fluctuating *Harmonics* (in a non-uniform manner)
4	Also, the presence of the Resonant frequencies at the Lyrics portion distinctly identify the regions with Lyrical-voicing in it
5	Harmonic activity is prominently observed in the frequency range above *3 KHz*, whereas in the prior range, the information about different types of sources can be obtained, like for Piano melody
7	Percussive sound effects can be distinctly highlighted by the SoE, leading towards the identification of the instrumental onsets, especially for Drums and Piano and some non-music sounds like Claps as well
8	SoE has also been observed to characterize high intensity vocal renditions
9	The Short-time spectrum reflects the excitation source information like polyphony, as singular vs. overlapped arrangement of the Harmonics

Table 3. Analysis of Singing Voice in Classical Music

S. #	*Key resulting observation*
1	Higher average F_0 is observed in the alaap regions as compared to the lyrics part for Female classical singers
2	For Male classical singing the regions with the lyrics part have higher F_0 value as opposed to alaap region
3	Energy is observed to be more in the voiced regions of the singing, i.e. lyrics parts, for the Female artists. Whereas, in male classical singing more modulation is present in the alaap regions making the energies more visible in alaap regions

(PDAs) like Autocorrelation and ZFF. Lyrical voicing manifests in the form of highly varying and non-uniform Harmonics. Resonances can also be insightful towards identifying the Lyrical-voicing regions. The deviation of ZFF based F_0 estimates can be helpful in identification of the regions with just the Music and Lyrical-Music Mixtures. The Harmonic peaks in the Short-time magnitude spectrums can be leveraged towards identifying excitation source information like the polyphony, sound sources, melody tracking, etc. SoE values are observed to distinctly characterize the onsets of various instruments like Drums, Piano, even the voices like Clap. SoE not only gives insights towards the chorus parts in the form of *high mean SoE*, but also for intense vocalizations.

The study on Classical Music was aimed at finding the distinct components in Indian ragas. Different Indian ragas from both Hindustani and Carnatic music were studied. Features studied were F_0 extraction using Autocorrelation of LP Residual and Zero Frequency Filtering and Signal Energy. F_0 value are higher in the alaap regions, the one with improvisation and lower in the regions with the more lyrics part. For instance, raga adana malhar and raga yamani hindol, F_0 values are significantly higher in case of female singers. However, in male classical singers, F_0 values are significantly higher in the lyrics region of the raga yaman malhaar. Energy was obtained to be higher for the lyrics region and lesser in case of alaap part.

The insights obtained towards music component onset identification can be helpful towards further exploratory work for characterizing music components like Accompaniments, Lyrical-voicing, Chorus, etc., annotation of the Music mixture in Time-Frequency Format and do the selective processing.

References

1. de Cheveigne, A.: Yin, a fundamental frequency estimator for speech and music. J. Acoust. Soc. Am. **111**(4), 1917–1930 (2002). doi:10.1121/1.1458024
2. Ephraim, Y., Malah, D.: Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator. IEEE Trans. Acoust. Speech Signal Process. **32**(6), 1109–1121 (1984)
3. Haykin, S.: An Introduction to Analog and Digital Communications. Wiley, New York (1989). http://www.loc.gov/catdir/toc/onix02/88015512.html
4. Li, Y., Wang, D.: Separation of singing voice from music accompaniment for monaural recordings. Trans. Audio, Speech Lang. Proc. **15**(4), 1475–1487 (2007). doi:10.1109/TASL.2006.889789
5. Martin, R.: Noise power spectral density estimation based on optimal smoothing and minimum statistics. IEEE Trans. Speech Audio Process. **9**(5), 504–512 (2001)
6. Mittal, V.K., Yegnanarayana, B.: Significance of aperiodicity in the pitch perception of expressive voices. In: INTERSPEECH 2014, 15th Annual Conference of the International Speech Communication Association, Singapore, 14–18 September, 2014, pp. 504–508 (2014). http://www.isca-speech.org/archive/interspeech_2014/i14_0504.html
7. Mittal, V.K., Yegnanarayana, B.: Study of characteristics of aperiodicity in Noh voices. J. Acoust. Soc. Am. **137**(6) (2015)

8. Ockelford, A.: Repetition in music: theoretical and metatheoretical perspectives. In: Royal Musical Association Monographs. Farnham, U.K., Ashgate (2005)
9. Oppenheim, A.V., Schafer, R.W., Buck, J.R.: Discrete-time signal processing, 2nd edn. Prentice-Hall Inc., Upper Saddle River (1999)
10. Ozerov, A., Philippe, P., Bimbot, F., Gribonval, R.: Adaptation of bayesian models for single-channel source separation and its application to voice/music separation in popular songs. IEEE Trans. Audio, Speech Lang. Process. **15**(5), 1564–1578 (2007). doi:10.1109/TASL.2007.899291
11. Rabiner, L., Juang, B.H.: Fundamentals of Speech Recognition. Prentice-Hall Inc., Upper Saddle River (1993)
12. Rafii, Z., Pardo, B.: Repeating pattern extraction technique (repet): a simple method for music/voice separation. IEEE Trans. Audio Speech Lang. Process. **21**(1), 73–84 (2013). doi:10.1109/TASL.2012.2213249
13. Rao, V., Ramakrishnan, S., Rao, P.: Singing voice detection in north indian classical music. In: Proceedings of the National Conference on Communications (NCC) (2008)
14. Sharma, S., Mittal, V.K.: Singing characterization using temporal and spectral features in indian musical notes. In: 2016 International Conference on Signal Processing and Communication. JIIT, Noida (2016)
15. Sharma, S., Mittal, V.K.: Window selection for accurate music source separation using repet. In: 2016 3rd International Conference on Signal Processing and Integrated Networks (SPIN), pp. 270–274 (2016). doi:10.1109/SPIN.2016.7566702
16. Sjölander, K., Beskow, J.: Wavesurfer-an open source speech tool
17. Sohn, J., Kim, N.S., Sung, W.: A statistical model-based voice activity detection. IEEE Signal Process. Lett. **6**(1), 1–3 (1999)
18. Yegnanarayana, B., Murty, K.S.R.: Event-based instantaneous fundamental frequency estimation from speech signals. IEEE Trans. Audio Speech Lang. Process. **17**(4), 614–624 (2009). doi:10.1109/TASL.2008.2012194

Design and Performance Analysis of Step Graded Dielectric Profile High Gain Flexible Textile Antennas for Radiolocation Military and Aeronautical Radio Navigation Applications

Kirtan Kaur[1], Sneh Kanwar Singh Sidhu[2], Aman Nag[3],
Raveena Bhatoa[2], and Ekambir Sidhu[4](✉)

[1] Department of Electronics and Communication Engineering,
Punjabi University, Patiala, India
[2] Department of Computer Engineering, Punjabi University, Patiala, India
[3] Department of Mechanical Engineering, Punjabi University, Patiala, India
[4] Department of Electrical and Computer Engineering,
University of Ottawa, Ontario, Canada
esidh097@uottawa.ca

Abstract. This paper emphasizes on the analysis of stacking different dielectric materials forming step graded profile along the thickness of the substrate in the microstrip patch antenna design. The different step graded dielectric profiles have been employed to design antennas and the performance of proposed antennas have been analyzed in terms of gain (dB), directivity (dBi), return loss (dB), VSWR, half power beam width (degrees), total efficiency (dB), side lobe level (dB) and impedance bandwidth (MHz). The stacking of three flexible textile materials namely fleece, felt and curtain cotton having dielectric constant of 1.04, 1.35 and 1.47, respectively have been used as substrate material to obtain six step graded dielectric profile antenna configurations which are P1, P2, P3, P4, P5 and P6. It has been concluded that the profile P3 having staircase dielectric profile is the best suited antenna configuration due to effective return loss, high gain, directivity and HPBW. The step graded dielectric profile antenna designs have been practically fabricated employing copper material as patch, feedline and ground having conductivity of 5.96×10^7 S/m. The performance of fabricated antennas has been analyzed practically by employing E5071C network analyzer and anechoic chamber. It has been observed that practical results intently match with the simulated results of the proposed antenna configurations. The proposed antenna configurations can be suitably employed for Military Radiolocation and Aeronautical Radio navigation applications as the proposed antenna configuration is resonant at 16.262 GHz which matches with the corresponding resonant frequency range of 15.6–16.6 GHz.

Keywords: Directivity · Flexible antenna · Gain · HPBW · Return loss · SLL · Step graded dielectric profile · Textile antenna · VSWR

S.M. Thampi et al. (eds.), *Advances in Signal Processing and Intelligent Recognition Systems*, Advances in Intelligent Systems and Computing 678,
DOI 10.1007/978-3-319-67934-1_7

1 Introduction

Microstrip patch antenna is a necessary and critical component of communication systems and a popular choice due to their ease of design, low profile and a compact structure. The important design aspect of microstrip antenna is that its bandwidth can be very easily varied by changing the shapes and thickness of substrate as well as by using multilayer substrate which is not possible in any other antenna [1]. Because of microstrip patch antenna's many unique and attractive properties there seems to be little doubt that it will continue to find many applications in the future. Its properties include, light weight, low profile, easy fabrication, compact and conformability to mounting structure [2, 3]. A simple microstrip patch antenna consists of conducting patch and ground plane with substrate of specific dielectric constant placed between them. The dimensions of microstrip patch antenna depends on the resonant frequency and value of dielectric constant of substrate. For good antenna performance, a thick dielectric substrate having a low dielectric constant is desirable since this provides better efficiency, larger bandwidth and better radiation [4]. The conventional patch antenna aches from a very serious intrinsic limitation of narrow bandwidth which has been thwarting its application in wide range of areas of wireless communication systems [5]. The multiple patches with one feeding patch and other parasitic patch result in multiple resonance frequencies [6]. A proper selection of patch geometry and gaps between the multiple patches can results in superimposing of adjacent frequency bands and thus providing large bandwidth [7].

The miniaturization can affect antenna efficiency characteristics like bandwidth (MHz), gain (dB), radiation efficiency (dB) and polarization purity [8]. The miniaturization approaches are based on either geometric manipulation (the use of bend forms, meandered lines, PIFA shape, varying distance between feeder and short plate, using fractal geometries [9–12].

The Defected Ground Structure (DGS) is one of the methods, which is used to miniaturize the size of microstrip antenna. The DGS is basically etching of a simple or complex shape in the ground plane for the better performance [13].

In this work, the microstrip patch antenna of substrate consisting of three different textile stacked materials having different dielectric constants have been designed and the effect of the stacking of dielectric materials in various antenna configurations on antenna parameters like gain, bandwidth, half power beam width, directivity, total efficiency and side lobe level have been observed. This research article covers various sections where proposed antenna geometries have been discussed in Sect. 2 while simulated results, experimental verification and conclusion have been discussed in Sects. 3, 4 and 5, respectively.

2 Antenna Geometry

The proposed step graded dielectric profile antennas have been designed and simulated using Computer Simulation Technology (CST) Microwave Studio 2016. The antenna has been fabricated using three different materials as substrate i.e. fleece, felt and curtain cotton having relative permittivity of 1.04, 1.35, 1.47 respectively and thickness

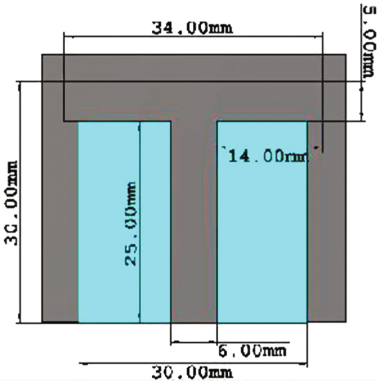


Fig. 1. Top view of the proposed antenna design

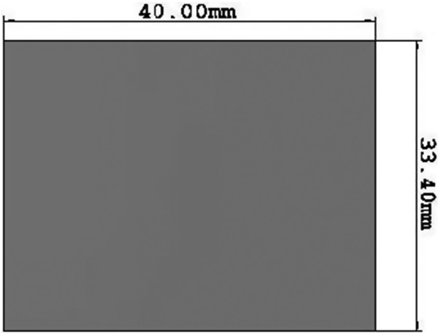


Fig. 2. Bottom view of the proposed antenna design

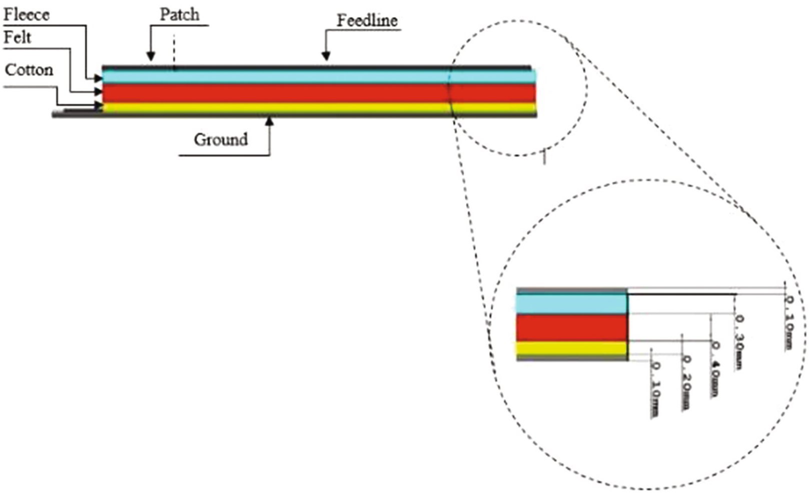


Fig. 3. Side view of the proposed antenna design

of 0.3 mm, 0.4 mm and 0.1 mm respectively as shown in Fig. 3. The proposed antenna design is compact size 40 × 33.4 mm^2. The Fig. 1 represents the top view of the proposed textile microstrip patch antenna in which the geometry of substrate and patch has been illustrated. The Fig. 2 shows the bottom view of the proposed antenna design in which shape and size of extended ground been illustrated. The Fig. 3 depicts the side view of the antenna design where thickness of various substrate has been shown. The copper material of thickness 0.1 mm having conductivity of 5.96 × 10^7 S/m has been used for radiating patch, feedline and ground.

The ground plane of the proposed antenna design has been extended to improve antenna bandwidth and return loss. The patch has been fed by a microstrip feed line of 6 mm width.

The feed line is adjusted so that the impedance of the antenna should closely match with the input impedance of SMA connector having impedance of 50 Ω. The dimensions of proposed antenna designs have been illustrated in Figs. 1, 2 and 3. The green color represents the fleece, red color indicates the felt and yellow color represents

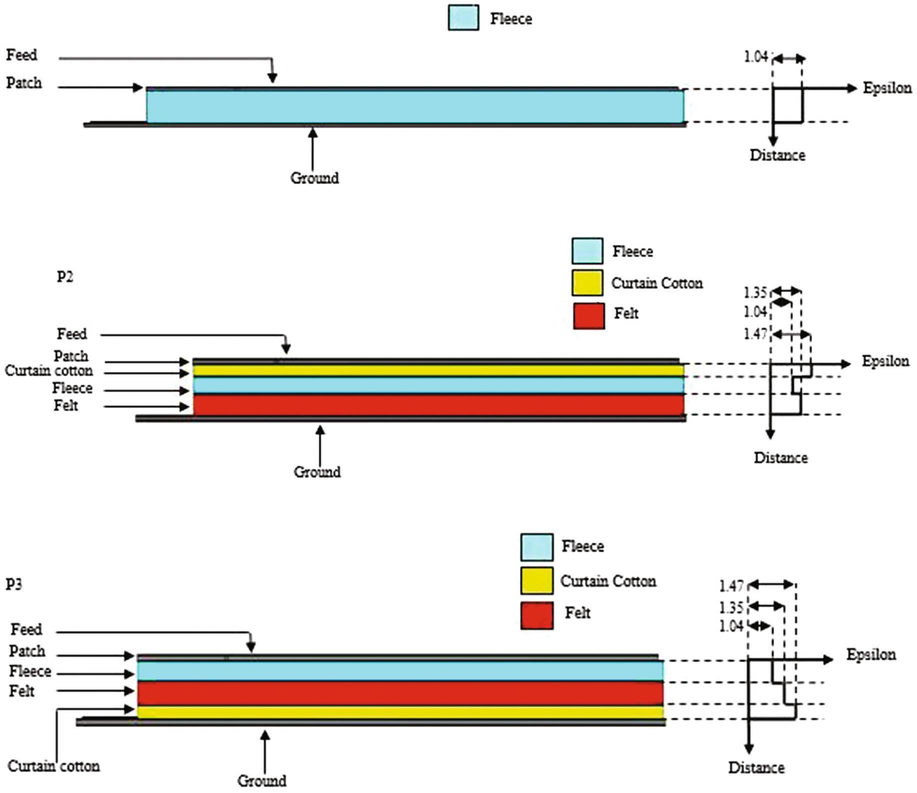


Fig. 4. (1) Side view of the proposed antenna configuration P1. (2) Side view of the proposed antenna configuration P2. (3) Side view of the proposed antenna configuration P3. (4) Side view of the proposed antenna configuration P4. (5) Side view of the proposed antenna configuration P5. (6) Side view of the proposed antenna configuration P6

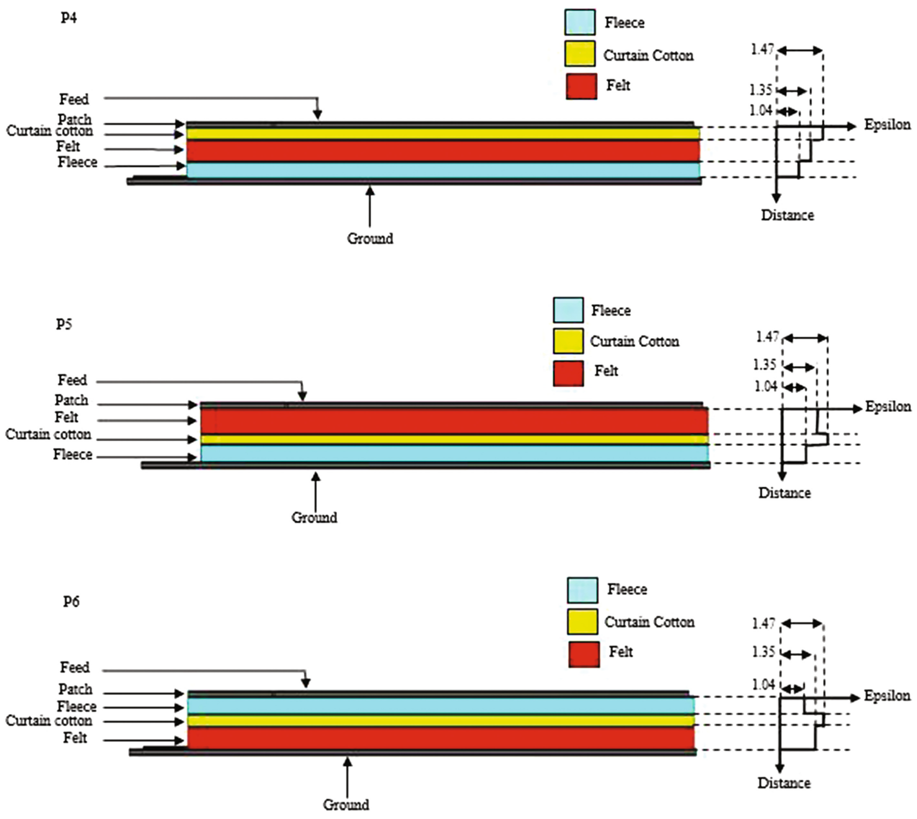


Fig. 4. (*continued*)

the curtain cotton in the Fig. 3. The stacking of three flexible textile materials have been done to obtain six step graded dielectric profile antenna configurations which have been designated as P1, P2, P3, P4, P5 and P6 this paper. The different profiles obtained by varying the position of stacked layers have been shown in Fig. 4.

3 Simulated Results

The novel concept of stacked substrates using three textile materials such as fleece, felt and curtain cotton have resulted in six profiles named P1, P2, P3, P4, P5, P6 which have been tested for impedance matching (ohms), return loss (dB), directivity (dBi), resonant frequency (GHz), VSWR, HPBW (degree) and their bandwidth (GHz) which can be elaborated from the Table 1 and Fig. 5. This table consists of all the simulated results of these profiles for the proposed antenna, where the red colored numerals indicates the highest value of the characteristic among the profiles while the green numerals indicates the lowest value of the characteristic among the profiles. This table ease our purpose to choose the best of the profile to be P3 as it has fascinating results

Table 1. Comparison of different step graded dielectric profile antennas

Profiles	Return loss (dB)	Gain (dB)	Directivity (dBi)	Bandwidth (MHz)	HPBW (degrees)	Efficiency (dB)
P1	−36.20	8.31	8.38	533	28.3	−0.064
P2	−34.64	8.54	8.66	444	29.5	−0.126
P3	−46.17	8.59	8.70	455	29.5	−0.103
P4	−33.45	8.54	8.65	434	29.5	−0.116
P5	−32.96	8.51	8.66	445	29.5	−0.148
P6	−41.19	8.48	8.60	496	28.7	−0.122
Optimum results	−46.17	8.59	8.7	533	28.3	−0.064

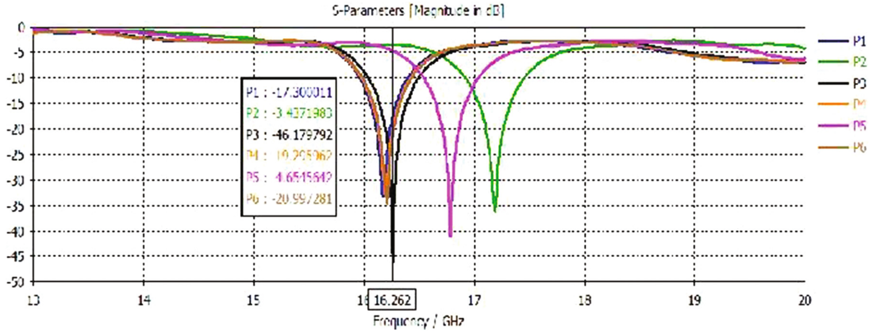


Fig. 5. Comparison of return loss plot of various step graded dielectric profile antenna configurations @ 16.26 GHz using CST Microwave Studio 2016

among the other profiles and with respect to P1 which is a single material substrate, specifically designed to distinguish our purpose of using different materials as a substrate. It can be comprehended from the results that P3 dishes us the prominent results

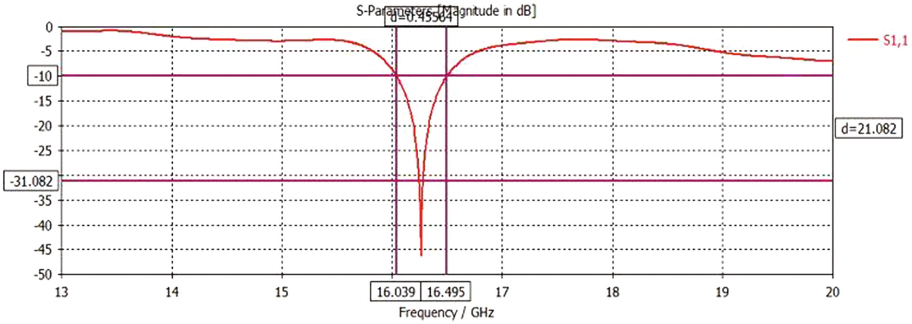


Fig. 6. Bandwidth of the simulated antenna design configuration P3 @ 16.26 GHz using CST Microwave Studio 2016

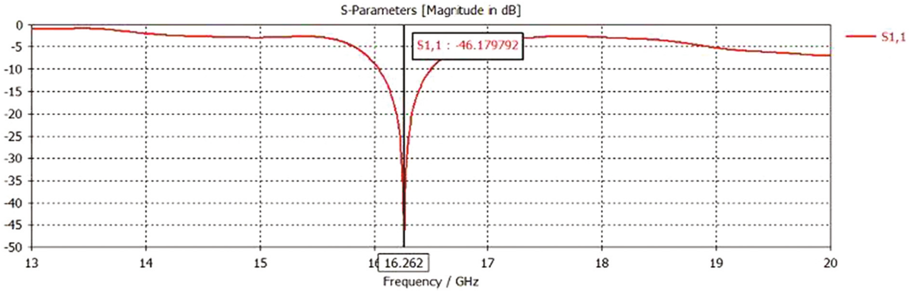


Fig. 7. Return loss plot of the simulated antenna design configuration P3 @ 16.26 GHz using CST Microwave Studio 2016

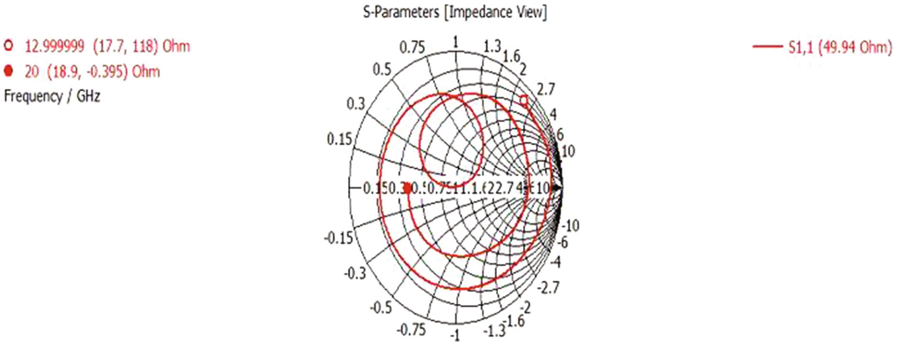
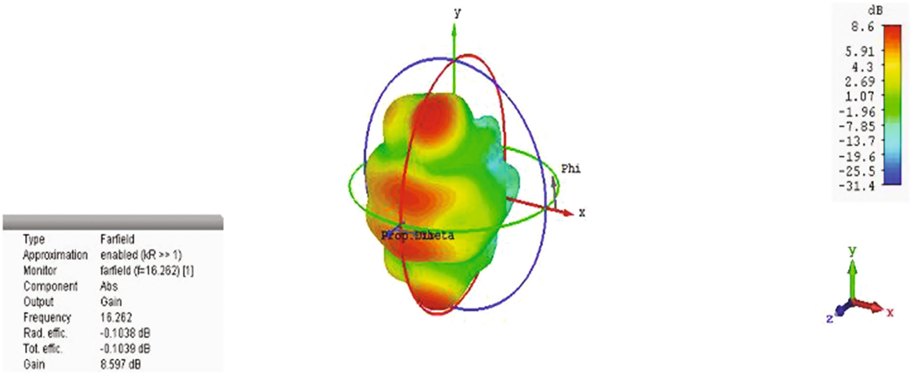
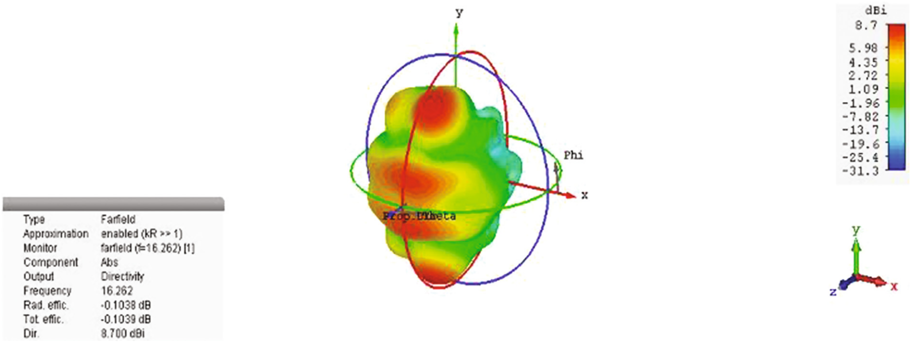


Fig. 8. Smith chart of the simulated antenna design configuration P3 @ 16.26 GHz using CST Microwave Studio 2016



Type	Farfield
Approximation	enabled (kR >> 1)
Monitor	farfield (f=16.262) [1]
Component	Abs
Output	Gain
Frequency	16.262
Rad. effic.	-0.1038 dB
Tot. effic.	-0.1039 dB
Gain	8.597 dB

Fig. 9. Gain of the simulated antenna design configuration P3 @ 16.26 GHz using CST Microwave Studio 2016



Type	Farfield
Approximation	enabled (kR >> 1)
Monitor	farfield (f=16.262) [1]
Component	Abs
Output	Directivity
Frequency	16.262
Rad. effic.	-0.1038 dB
Tot. effic.	-0.1039 dB
Dir.	8.700 dBi

Fig. 10. Directivity of the simulated antenna design configuration P3 @ 16.26 GHz using CST Microwave Studio 2016

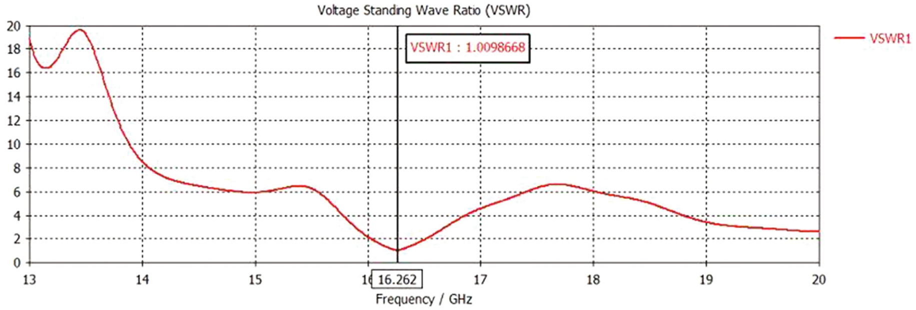
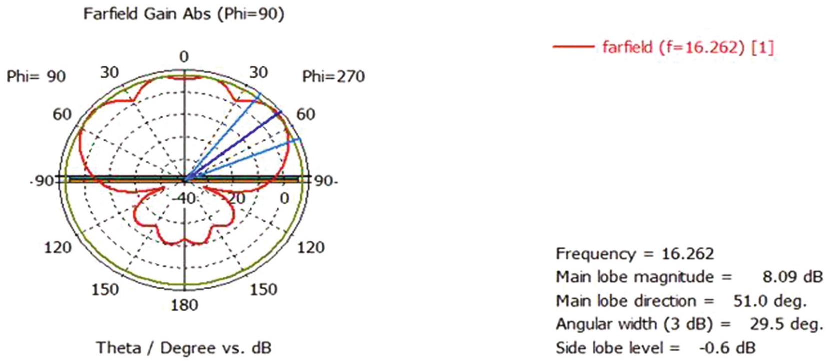


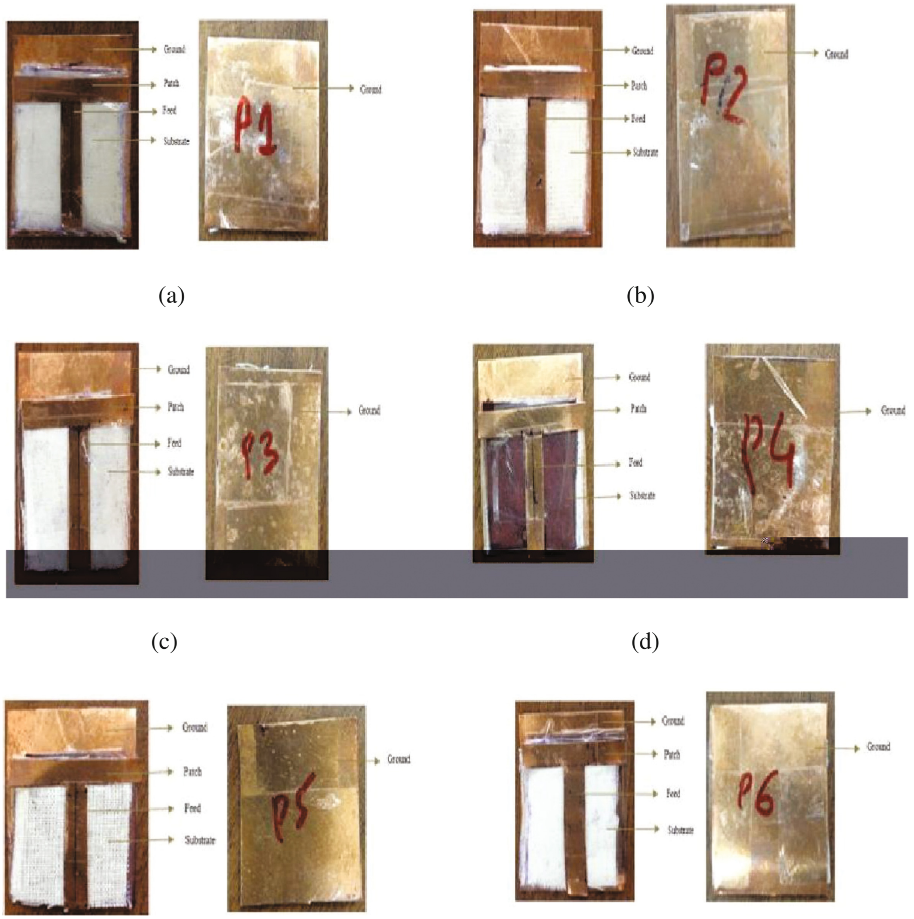
Fig. 11. VSWR plot of the simulated antenna design configuration P3 @ 16.26 GHz using CST Microwave Studio 2016



Frequency = 16.262
Main lobe magnitude = 8.09 dB
Main lobe direction = 51.0 deg.
Angular width (3 dB) = 29.5 deg.
Side lobe level = -0.6 dB

Fig. 12. Half power beam width (HPBW) plot of the simulated antenna design configuration P3 @ 16.26 GHz using CST Microwave Studio 2016

in majority of our parameters as can be analyzed that it is resonant at the frequency of 16.262 GHz covering a bandwidth of 455.69 MHz in the frequency range of 16.039–16.495 GHz offering the return loss of −46.179792 dB as can be seen from the simulated results plot in Figs. 6 and 7. It can also be analyzed that it provides the prominent high gain of 8.597 dB and directivity of 8.7 dBi at the resonant frequency of 16.262 GHz. It has been perceived that the impedance of the designed antenna is 49.94 Ω which closely matches the desired SMA port impedance of 50 Ω as shown in Fig. 8 thus ensuring maximum power transfer from port to antenna and vice versa. The gain and directivity plots of the proposed antenna can be viewed in Figs. 9 and 10. The VSWR plot in Fig. 11 also indicates its proficiency as it lies in the maximum



(a) (b)

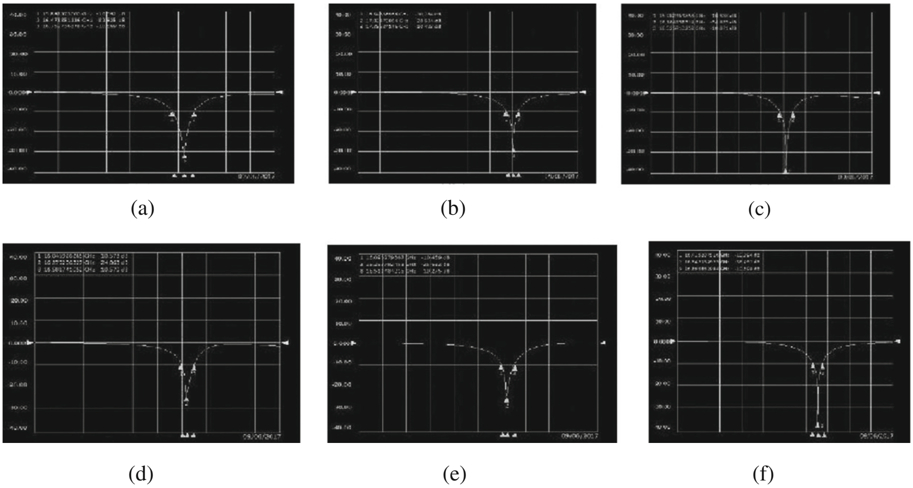
(c) (d)

(e) (f)

Fig. 13. (a–f) Top and bottom view of the step graded dielectric profile fabricated antenna configuration P1, P2, P3, P4, P5 and P6, respectively acceptable range of less than 2 at the resonant frequency. The half power beam width (HPBW) plot in Fig. 12 indicates that the designed antenna has HPBW of 29.5°. The proposed step graded dielectric profile antenna design P3 can be effectively employed for radiolocation (ECA36), radiolocation (military) and radio navigation applications (15.6–16.6 GHz).

4 Experimental Verification

The prototype of various step graded dielectric profile antenna have been fabricated and tested by deploying E5071C network analyzer and anechoic chamber. The top view and bottom view of the various contrived antennas along with their dielectric constant patterns is shown in the Fig. 13 below. Although the tested and simulated results show reasonable agreement through the entire band but the little discrepancy has been observed due to feed point radiations and reflection losses. The practical results of the fabricated antenna designs have been shown in the Fig. 14.



(a) (b) (c)

(d) (e) (f)

Fig. 14. (a–f) Return loss plot of step graded dielectric profile fabricated antenna configuration P1, P2, P3, P4, P5 and P6, respectively

5 Conclusion

In this work, a flexible step graded dielectric microstrip patch antenna has been designed and simulated using CST microwave studio 2016. The stacking of three textile materials namely fleece, felt and curtain cotton having dielectric constant of 1.04, 1.35 and 1.47 respectively have been used as substrate material. The aim was targeted at observing the variations in the microstrip patch antenna characteristics which are gain (dB), directivity (dBi), return loss (dB), half power beam width (degrees), total efficiency (dB), side lobe level (dB) and impedance bandwidth (MHz) by varying the stacked substrate layers of three different textile materials. The six different configurations have been designed and analyzed by varying the position of three stacked layers in the substrate. The performance analysis of six configurations have been obtained by varying positions of felt, fleece and curtain cotton used as substrate materials. It has been concluded that profile P3 is the best suited antenna configuration due to effective return loss, high gain, directivity and HPBW. The proposed antenna configurations have been designed for Military radiolocation applications and for radiolocation (ECA36) applications.

Acknowledgement. My deep gratitude goes to Assistant Professor Ekambir Sidhu, who expertly guided me through the research work and his direction helped me to form the structure of this research project. His unwavering enthusiasm for electronics steered our way throughout and we thank him for the guidance and at the last but not the least my appreciation also extends to our colleagues to keep a constant check on our progress.

References

1. Balanis, C.A.: Antenna Theory Analysis and Design, 2nd edn. Wiley, USA (2004)
2. Adegoke, O.M., Eltoum, I.S.: Analysis and design of rectangular microstrip patch antenna at 2.4 GHz WLAN applications. Int. J. Eng. Res. Technol. (IJERT). **3**(8) (2014)
3. James, J.R., Hall, P.S.: Handbook of Microstrip Antennas. IEEE Electromagnetic wave series, vol. 28. Peter Pergrinus, London (1989)
4. Constantine, A.B.: Antennas Theory—Analysis and Design, 3rd edn. Wiley, USA (1997)
5. Roy, S.S., Naresh, K.M., Saha, C.: Resistively loaded slotted microstrip patch antenna with controllable bandwidth. In: International Symposium on Antennas and Propagation (APSYM) (2016)
6. Pozar, D.M.: Microstrip Antennas. Proc. IEEE Trans. **80**, 79–91 (1992)
7. Garg, R., Bahl, I., Bhartia, P.: Microstrip Antenna Design Handbook. Artech House, Boston (2001)
8. Tarbouch, M., El Amri, A., Terchoune, H.: Compact CPW-Fed Microstrip Octagonal patch antenna with H slot for WLAN and WIMAX Applications (2017)
9. Chen, H.-T., Wong, K.-L., Chiou, T.-W.: Pifa with a meandered and folded patch for the dual-band mobile phone application. IEEE Trans. Antennas Propag. **51**(9), 2468–2471 (2003)
10. Reha, A., El Amri, A., Benhmammouch, O., Oulad Said, A.: Fractal antennas: a novel miniaturization technique for wireless networks. Trans. Netw. Commun. **2**(5), 20–36 (2014)
11. Sun, S., Zhu, L.: Miniaturised patch hybrid couplers using asymmetrically loaded cross slots. IET Microwave Antennas Propag. **4**(9), 1427 (2010)
12. Chi, P.-L., Waterhouse, R., Itoh, T.: Antenna miniaturization using slow wave enhancement factor from loaded transmission line models. IEEE Trans. Antennas Propag. **59**(1), 48–57 (2011)
13. Chater, N., Mazri, T., Benbrahim, M.: Design and Simulation of Microstrip Patch Array Antenna for Electronic Scanning Radar Application (2017)
14. http://www.erodocdb.dk/docs/doc98/official/pdf/ERCRep025.pdf

An Experimental Setup of DSA Algorithm Suitable for High Bandwidth Data Transfer Using USRP and GNU Radio Companion

S. Naveen Naik(✉) and B. Malarkodi(✉)

Department of Electronics and Communication Engineering,
NIT Trichy, Trichy, Tamilnadu, India
naveen.iisc2@gmail.com, malark@nitt.edu

Abstract. In this paper we present the experimental implementation of dynamic spectrum access (DSA) algorithm using universal software radio peripheral (USRP) and GNU Radio. The setup contains two primary users and two cognitive radios or secondary users. One primary user is fixed and the other is allowed to change its position randomly. Depending upon the position of the primary user the cognitive user will use the spectrum band where the detected energy is below certain predefined threshold level. The cognitive radio users are also programmed to operate independently without interfering with each other using energy detection algorithm for spectrum sensing. The modulation scheme is set to GMSK for secondary user performing data transmission. This experimental setup is used to analyze the quality of video transmission using DSA which provides the insight regarding the possibility of using free spectrum space to improve the performance of the system and its advantage over a non-DSA system. From the experiment it is shown that under congestion and interference DSA perform better than a non- DSA system.

Keywords: Cognitive radio · Dynamic spectrum access · IEEE 802.22 · Energy detection · USRP · GNU radio

1 Introduction

The current static allocation of the frequency spectrum is not efficient as studies by the FCCs Spectrum Policy Task Force (SPTF) reported vast temporal and geographic variations in the usage of allocated spectrum with utilization ranging from 15% to 85% [1]. These measurements seriously question the efficiency of the current regulatory policies and spectrum allocation. Furthermore with the increasing proliferation of wireless devices, the spectrum is becoming crowded and allocation of new spectrum is not scaling proportionally to meet the increasing user demands. This has forced researchers to explore new technologies to mitigate this problem. One of the technologies that are actively under research to increase the capacity of wireless system is dynamic spectrum access (DSA) which aims at improving the utilization of crowded otherwise underutilized spectrum in time, frequency and space. With DSA, the unlicensed user or secondary user (SU) can access and utilize the available spectrum when it is not being used by the licensed users. It also allows the SU to move to another free

S.M. Thampi et al. (eds.), *Advances in Signal Processing and Intelligent Recognition Systems*, Advances in Intelligent Systems and Computing 678,
DOI 10.1007/978-3-319-67934-1_8

spectrum whenever the licensed user appears to reuse the specified spectrum band where SU was operating.

In order to realize DSA, a number of available technologies can be used for instance using adaptive radio [2], cognitive radio (CR) [3], and reconfigurable radio [4]. There are certain requirements that need to be fulfilled such as sensing the spectral environment over a wide bandwidth. Co-exists with licensed users and do not interfere with them and adapt the operating parameters to their environment. The device which fulfils all such requirement perfectly is cognitive radio (CR). It is a fundamental requirement that CR user or secondary users, which are the unlicensed users, must detect the spectrum hole efficiently to avoid interference to the primary user (PU) and exploit the spectrum holes for required throughput and quality-of service (QoS). In order to provide the features like spectrum sensing and spectrum mobility, the unlicensed user must have a very flexible radio transceiver. The most appropriate means to realize this radio is through the use of software defined radio (SDR). Towards this goal, recent research projects have produced a number of viable SDR platforms for research experimentation and deployment, such as KNOWS [5], SORA [6], USRP GNU Radio [7], and WARP [8].

In this paper we use USRP and GNU Radio for the implementation of DSA. USRP act as the hardware component (RF front end) that is used to receive the signal in the environment and then transmit the received information to GNU Radio, which is the software part, for further signal processing. One appealing feature of the GNU Radio is that the transceiver modules are defined by using software, thus offering great flexibility to its users. The aim of this paper is to analyze the feasibility of DSA technique using GNU radio and USRP and use it for video transmission to see whether the DSA technique is suitable for high bandwidth data transfer, such as video, or not and compare it with non- DSA system.

This paper is organized as follows. Section 2 describes the free spectrum detection mechanism using energy detection.

Section 3 explains the operation of DSA system implemented in the system. The experimental setup is presented in Sect. 4 followed by performance evaluation and conclusion in Sects. 5 and 6 respectively.

2 Spectrum Sensing Model

Before the unlicensed user can use the spectrum belonging to the licensed user it needs to identify the spectrum where the primary user is not present. This is done by spectrum sensing.

Spectrum sensing is the art of performing measurements on a part of the spectrum and forming a decision related to spectrum usage based upon the measured data. In this experimental setup we use energy detection for identifying the free spectrum as it is simple to implement and provide satisfactory results. Typically, local sensing for primary signal detection is formulated as a binary hypothesis problem as follows:

$$x_i(t) = \begin{cases} n_i(t),\ H_0 \\ h_i(t) \cdot n_i(t), H_1 \end{cases} \tag{1}$$

where $x_i(t)$ denotes the received signal at the CR user, $s_i(t)$ is the transmitted PU signal, $h_i(t)$ is the channel gain of the sensing channel, $n_i(t)$ is the zero-mean additive white Gaussian noise (AWGN), H_0 and H_1 denote the hypothesis of the absence and the presence, respectively, of the PU signal in the frequency band of interest. In the energy detection approach the radio frequency energy in the channel or the received signal strength indicator (RSSI) is measured over a fixed bandwidth W over an observation time window T to determine whether the PU is active in the frequency band of interest.

The diagram of CR receiver implemented using USRP and GNU Radio is shown in Fig. 1. Specifically, spectrum sensing is performed using energy detection method. High sample rate processing like digital up and down conversion is performed in the FPGA, while rest of signal processing is implemented in C++/Python on the host machine. SBX daughterboard fetches the RF signal from the environment and converts it to Intermediate Frequency (IF). After converting it to IF the signal is passed to ADC, USRP-N210 contains two 14-bit ADC which provides sampling rate of 100 MS/s [9]. The ADC after sampling and digitizing passes the data to FPGA. The main task for the FPGA is to down convert the remaining frequency and data rate conversion. After processing, FPGA transfers the results to gigabit Ethernet controller which passes it over to the host computer where the rest of the signal processing tasks are performed.

Some Common Mistakes. The stream vector block receives the stream from Ethernet port for further processing. Its output is sent through an FFT block and its averaged output is compared to a predetermined threshold. Then, the decision block determines the absence or presence of PUs based on the average of FFT samples being less than or greater then a predetermined threshold.

For the purpose of spectrum sensing we use the script *DSA_spectrum_sence.py.* This script is a modified version of the script *usrp_spectrum_sense.py* that comes with the GNU Radio so that it can operate with the USRP N200 and serves our need. The USRP N200 is capable of scanning bandwidth of 25 MHz as opposed to 8 MHz on USRP1. Since it scans a larger bandwidth in a given time it can detect the spectrum opportunity fast and results in increased spectrum utilization.

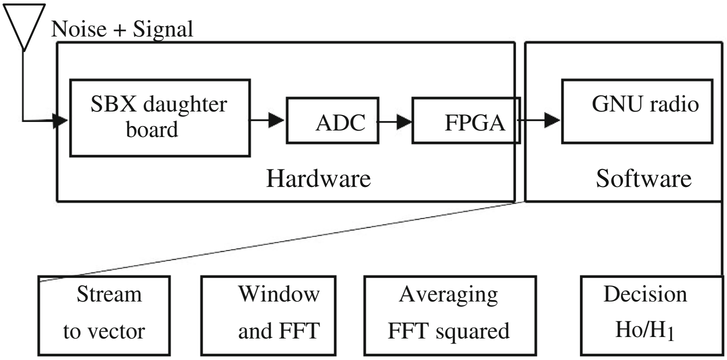


Fig. 1. Schematic diagram of CR receiver with USRP and GNU Radio using energy detection method for spectrum sensing

3 Spectrum Sharing

Spectrum sharing allows the unlicensed user to access the spectrum that belongs to the licensed spectrum which is dedicated to a specific technology. There are two types of spectrum sharing known as underlay and overlay spectrum sharing respectively. In underlay spectrum sharing, unlicensed user is strictly limited in emitting signal such that it has to be below the designated threshold. This type of spectrum sharing allows SU to simultaneously share the spectrum with licensed user in all dimensions which is time, frequency and space. In contrast, the overlay spectrum sharing disallows the unlicensed user to simultaneously use the same frequency which is in used by the licensed user due to the interference that it may cause. In this experiment we use overlay spectrum sharing for implementing the DSA for video transmission. Figure 2 shows the simulation of overlay spectrum sharing for DSA in MATLAB. The figure plots the power spectral density of the primary and secondary transmitter. It can be seen that the secondary user utilizes only those frequency (shown on the right side of the figure) that are not in use by the primary user.

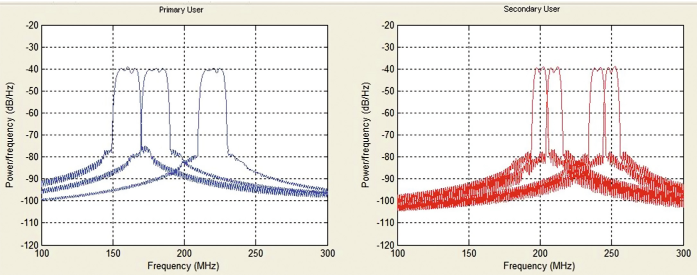


Fig. 2. MATLAB simulation of DSA using OFDM.

4 Experimental Setup

In this experimental setup, we demonstrate the DSA using a USRP/GNU Radio test bed shown in Fig. 3. The frequency band of operation is 2.3 GHz to 2.42 GHz. The test bed consist of two SU and two PU. To model two PU and 2 SU four USRP are required to model each user. Another extra USRP is used to receive the video transmission by secondary user. The description of the system setup for experimental setup is as follows:

1. *Primary Users*
   There are two primary user transmitters in this experimental setup. The first primary transmitter transmits a narrowband signal which changes its frequency randomly after ten second. It sweeps the specified frequency range except 2.4 GHz. The other primary transmitter is fixed at 2.4 GHz and uses OFDM modulation.

2. *Secondary Users*

   This experiment uses two SU. The first SU transmits video signal and the second SU transmit OFDM signal of randomly generated bits. We use one more USRP to capture the video transmission. SU consists of a transceiver and performs following function for DSA:

   (a) *Spectrum Sensing*

   This is used to observe the RF spectrum of the primary and the secondary systems. This is done by the receiver in the secondary user that continuously scans the entire spectrum of interest and informs the transmitter about the frequency that is below a certain threshold and ask it switch to that frequency. This switching is only done when the energy sensed in that band is less than the energy sensed in band where the transmitter is currently operating. This ensures there is no unnecessary switching.

   (b) *Transmission*

   The secondary user starts their transmission at predefined frequency. After that it switches the operating frequency as soon as it finds a new vacant spectrum band.

   (c) One more extra USRP is used for the reception of the video transmission. The synchronization is done by connecting to the same PC as used by the secondary user transmitting video signal. The video receiver reads from the file where the spectrum sensing stores the free frequency bands. This allows the receiver to follow the transmitter frequency.

The device used in this experiment is Ettus Research USRP N200 which connects to the PC via a Gigabit Ethernet, Laptops, Spectrum analyzer. The daughterboard for RF frontend is SBX board capable of operating in the range of 400 MHz–4.4 GHz. SBX daughterboard is capable for both transmitting and receiving. All USRP boards are connected to individual laptops.

Initially, the fixed primary user is turned on and the secondary user senses the spectrum and transmits at the predetermined frequency. Another secondary user transmits video stream and changes frequency of operation according to the spectrum sensing result. Both secondary users coexist without interfering with each other and the primary users. Next we describe the mechanism that enables the coexistence and DSA.

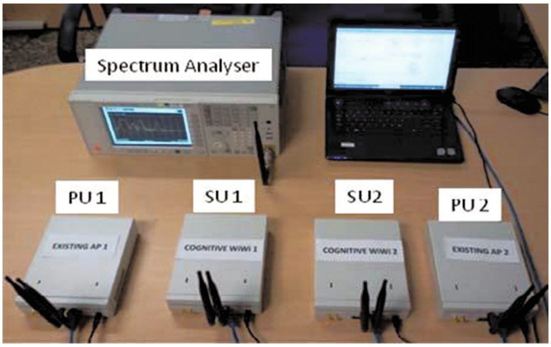


Fig. 3. Experimental setup for dynamic spectrum access

5 System Operation

We used Linux operating system Ubuntu 10.04 to run GNU Radio. The proposed design of the CR system for DSA is illustrated by the block diagram in Fig. 4. The video stream for transmission is produced by VLC player using H.264 encoding. The encoded stream is then sent to a UDP port. This UDP port is used to connect the output of VLC player with the GNU Radio. The GNU Radio to listen to that specified port which then performs all the necessary operation required for the transmission of the encoded video stream. The output of GNU Radio is connected to USRP which then transmit the data received from GNU Radio after performing digital to analog conversion.

Initially the transmission starts at predefined frequency which is set to 2.4 GHz, which then changes according the feedback given from the spectrum decision block.

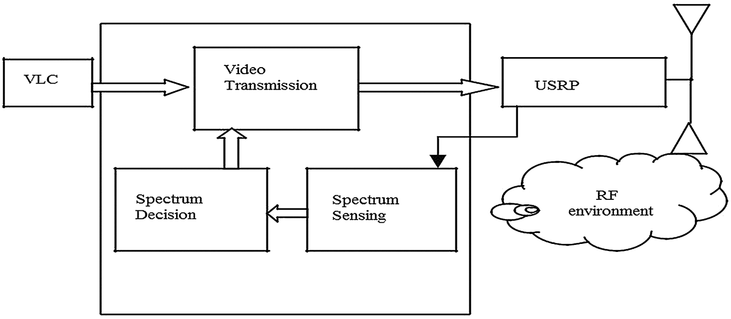
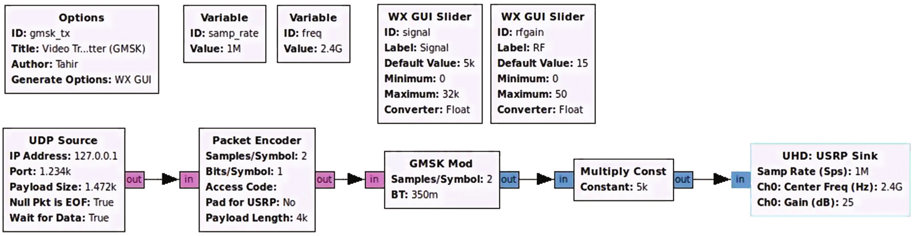


Fig. 4. System diagram for video transmission with DSA implemented using GNU Radio

The spectrum sensing scans the RF environment and determines those frequencies which are below a certain threshold. These frequencies are then sent to spectrum decision block which decide on whether to change the current operational frequency or not. By default, the GNU Radio instructs the USRP to jump to the first detected frequency that is below threshold. Then after that only those frequencies are selected for operation whose measured energy is below the last frequency of operation. In the following subsection description of various parts is given that realizes DSA for video transmission.

A. *Video Transmission*

The block diagram for video transmission is shown in Fig. 5. This block diagram is generated using GNU Radio companion (GRC) and is called flow graph. The first block in the flow graph is a UDP source that receives the video input stream which is encoded in H.264 format by VLC player. VLC player sends the video stream to UDP port 1234. After the encoded video stream is received it is passed to encoder to be encoded in packets. Here we use 1bit/symbol and number of samples per symbol to be two. After forming packets the packetized data is modulated using GMSK modulation followed by multiplying the signal in order to boost the



Options
ID: gmsk_tx
Title: Video Tr...tter (GMSK)
Author: Tahir
Generate Options: WX GUI

Variable
ID: samp_rate
Value: 1M

Variable
ID: freq
Value: 2.4G

WX GUI Slider
ID: signal
Label: Signal
Default Value: 5k
Minimum: 0
Maximum: 32k
Converter: Float

WX GUI Slider
ID: rfgain
Label: RF
Default Value: 15
Minimum: 0
Maximum: 50
Converter: Float

UDP Source
IP Address: 127.0.0.1
Port: 1.234k
Payload Size: 1.472k
Null Pkt is EOF: True
Wait for Data: True

Packet Encoder
Samples/Symbol: 2
Bits/Symbol: 1
Access Code:
Pad for USRP: No
Payload Length: 4k

GMSK Mod
Samples/Symbol: 2
BT: 350m

Multiply Const
Constant: 5k

UHD: USRP Sink
Samp Rate (Sps): 1M
Ch0: Center Freq (Hz): 2.4G
Ch0: Gain (dB): 25

Fig. 5. Flow graph for video transmission in the experimental setup.

amplitude of modulated signal so it can be easily detected. After this the samples are sent out to USRP through Ethernet port. The USRP after performing digital to analog conversion sends to the SBX daughter board for transmission in the RF environment. All the blocks used in the building this flowgraph comes as standard blocks within GNU Radio Companion. However it possible to built custom blocks to suit particular need.

B. *Video Reception*

The flow graph generated in GRC for video reception is shown in Fig. 6. The first block in video reception is the USRP source. This USRP block provide interface to GNU radio via Ethernet port. The USRP captures the transmitted data from the environment and converts in a form suitable for processing by the laptop. The captured data is demodulated using GMSK demodulator and forwarded to the packet decoder which performs the inverse operation of the packet encoder to generate a bit stream. This bit stream is forwarded to UDP port so that it can be sent to VLC player displaying the video.

C. *Spectrum Sensing*

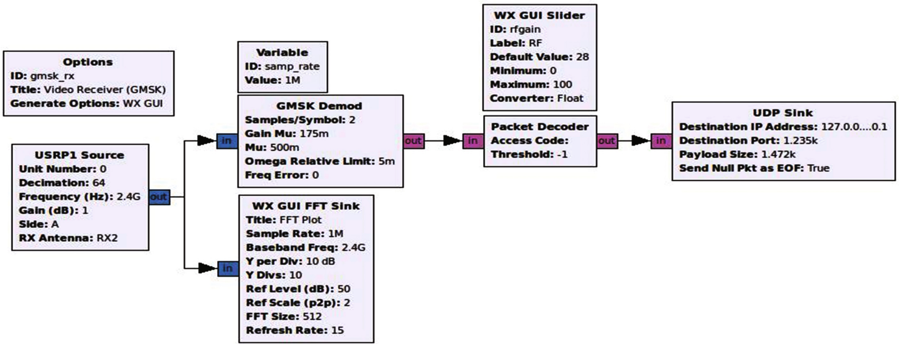
Spectrum sensing is started in secondary user by running the script *DSA_spectrum_sence.py* in terminal. This function instructs the receiver part of the secondary user to perform spectrum sensing continuously between 2.3 GHz to 2.42 GHz. The following command is used in the terminal,

$$./\mathit{DSA\_spectrum\_sence.py}\ 2.3G\ 2.42G$$

During this process it identifies the portion of the spectrum which is below the specified threshold and stores in the file.

D. *Secondary user operation*

The secondary users start their operation at 2.4 GHz. The first secondary user is turned and it operates on 2.4 GHz until it finds a vacant and switches to it. The second secondary user is turned on only when first secondary user switches the frequency of operation from 2.4 GHz. This is done to avoid interference among the secondary users.



Options
ID: gmsk_rx
Title: Video Receiver (GMSK)
Generate Options: WX GUI

USRP1 Source
Unit Number: 0
Decimation: 64
Frequency (Hz): 2.4G
Gain (dB): 1
Side: A
RX Antenna: RX2

Variable
ID: samp_rate
Value: 1M

GMSK Demod
Samples/Symbol: 2
Gain Mu: 175m
Mu: 500m
Omega Relative Limit: 5m
Freq Error: 0

WX GUI FFT Sink
Title: FFT Plot
Sample Rate: 1M
Baseband Freq: 2.4G
Y per Div: 10 dB
Y Divs: 10
Ref Level (dB): 50
Ref Scale (p2p): 2
FFT Size: 512
Refresh Rate: 15

WX GUI Slider
ID: rfgain
Label: RF
Default Value: 28
Minimum: 0
Maximum: 100
Converter: Float

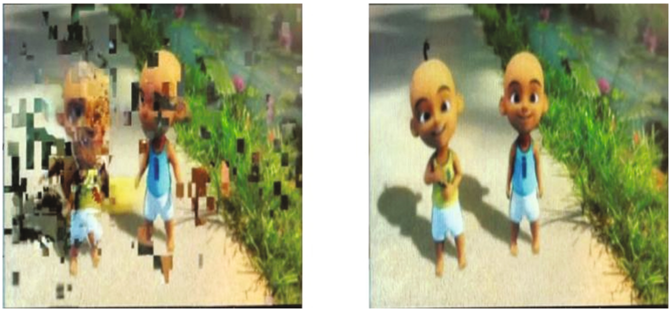
Packet Decoder
Access Code:
Threshold: -1

UDP Sink
Destination IP Address: 127.0.0....0.1
Destination Port: 1.235k
Payload Size: 1.472k
Send Null Pkt as EOF: True

Fig. 6. Flow graph for video reception in the experimental setup.

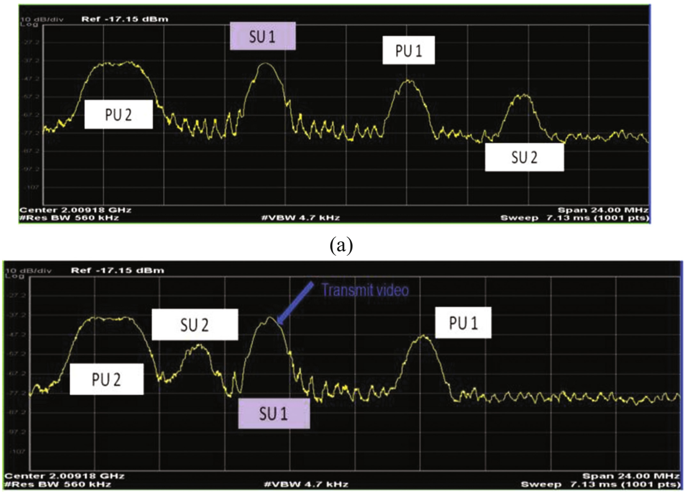
6 Performance Evaluation

In order to evaluate the performance of video transmission under dynamic spectrum access, we conducted experiments based on the setup described in previous section with and without dynamic spectrum access. As seen from the Fig. 7(a) under the case of no DSA due to interference in the overcrowded band and the limited bandwidth, the video quality degrades which is evident from the bad pixel appeared in the received video. The operation continues in the same frequency band in which it started (i.e. 2.4 GHz). Even if some interferer comes in and starts transmitting there is no mechanism to find new spectrum band that may be free of interference. On the other hand with DSA the secondary user who initially starts transmitting at 2.4 GHz changes frequency as soon as it finds a new frequency with measured energy below threshold and avoids all the bands where the measured energy higher than the threshold. Due to this the video quality is not degraded and the picture quality is much clearer without any disruptions as seen from the Fig. 7(b). Figure 8 shows the spectrum occupancy of the all the users in the 2.3 GHz to 2.42 GHz. From the spectrum it can be seen that all the users coexist without interfering with each other.



(a) (b)

Fig. 7. Snapshot of video sequence for (a) without DSA (b) with DSA



(a)

(b)

Fig. 8. Spectrum analyzer showing snapshot of spectrum occupancy when all the PU and SU are transmitting

7 Conclusion

In this paper we have presented an experimental setup to analyze the feasibility of using DSA to improve the performance of application like video transmission. From the experimental implementation it was shown that the system using DSA perform better than the one with non-DSA. The video quality in the DSA case was superior to the non-DSA case. This is due to the inability of non-DSA device unused band that is free and has less congestion. Due to frequency changing in the DSA setup there were some packet loss due to the inability of spontaneous synchronization between transmitter and receiver frequency. Although, there were some packet loss the performance degradation was not high enough to cause visible distortion in the received data. However, this was very simple lab scale setup to highlight the importance of DSA. The future work for DSA will consider the use of TV white space and more complex algorithm with cooperation among secondary users to improve the performance and make it suitable for real time services.

References

1. Spectrum policy task force report. In: Federal Communications Commission, Technical report 02-155, November 2002
2. International Telecommunication Union (ITU). Handbook frequency adaptive communication systems and networks in the MF/HF bands edn. (2002)
3. Mitola III, J.: Cognitive radio: An integrated agent architecture for software defined Radio, Ph.D. thesis, KTH- Royal Institute of Technology, Stockholm, Sweden (2000)

4. Baldini, G., et. al.: Reconfigurable radio systems for public safety based on low-cost platforms. In: Euro ISI 2008. LNCS, vol. 5376, pp. 237–247 (2008)
5. Chandra, R., et al.: A case for adapting channel width in wireless networks. In: Proceedings of SIGCOMM (2008) Conference on Data Communication, vol. 38, issue 4, October 2008
6. Tan, K.: SORA: High performance software radio using general purpose multi-core processors. Mag. Commun. ACM **54**(1), 99–107 (2011)
7. GnuRadio. http://gnuradio.org/redmine/projects/gnuradio/wiki
8. Wireless open-access research platform. http://warp.rice.edu/
9. Ettus Research. https://www.ettus.com/product

Influence of Filter Bank Structure on the Statistical Significance of Coefficients in Cepstral Analysis for Acoustic Signals

Akhil Jose[1](✉), Justin Joseph[2], Glan Devadhas[3], and M.M. Shinu[3]

[1] Applied Electronics and Instrumentation Department, ASIET, Kalady, India
akhilvjose@gmail.com
[2] Applied Electronics and Instrumentation Department, SJCET, Palai, India
[3] Electronics and Instrumentation Department, VJEC, Kannur, India

Abstract. Several programmed speech recognition and classification methods use Mel Frequency Cepstral Coefficients (MFCC) features. This paper presented a modification of Mel filter bank structure to a linear filter bank structure and obtained coefficients are named as Linear Frequency Cepstral Coefficients (LFCC). To study about the class discriminability of MFCC and LFCC, analysis of statistical significance of both the features is carried on specimens of two dissimilar databases. The first database is low-frequency heart sound signals of different classes and second is high-frequency music signals of various genres. Four different feature sets are formed by performing MFCC and LFCC feature extraction on the two databases. Kruskal-Wallis test is conducted on all four feature sets to investigate the inter-class variability and intra-class similarity of the features extracted. Further, box plots are plotted and analysed for better appreciation of feature sets and its ability to classify acoustic signals.

Keywords: Audio features · Filter bank · Cepstrum · MFCC · LFCC · Heart sounds · Music genre · Statistical significance

1 Introduction

Please Mel Frequency Cepstral Coefficients (MFCC) is extensively used features in audio signal processing. These coefficients as a group make up a Mel-Frequency Cepstrum (MFC) based on linear cosine transform of log power spectrum on nonlinear mel scale of frequency. MFC is a depiction of the short-term power spectrum of an acoustic signal. The nonlinear mel scale in MFC estimates the response of human auditory system more accurately than linearly spaced normal cepstrum.

MFCC has been in use for a long time now [1, 2] and is the most widely used speech feature in audio signal processing [3]. Few of the most recent applications include recognition of speech [4–6] and emotion [7, 8], action classification [9] and biomedical signal processing [10–12]. Many researchers have tried to improve the significance of MFCC features by modifying it in many ways. In this paper, the modification in the filter bank structure and its effect on the statistical significance is discussed. Also, an evaluation of the statistical significance of the Linear Frequency Cepstral Coefficients (LFCC) and the MFCC is conducted.

S.M. Thampi et al. (eds.), *Advances in Signal Processing and Intelligent Recognition Systems*, Advances in Intelligent Systems and Computing 678,
DOI 10.1007/978-3-319-67934-1_9

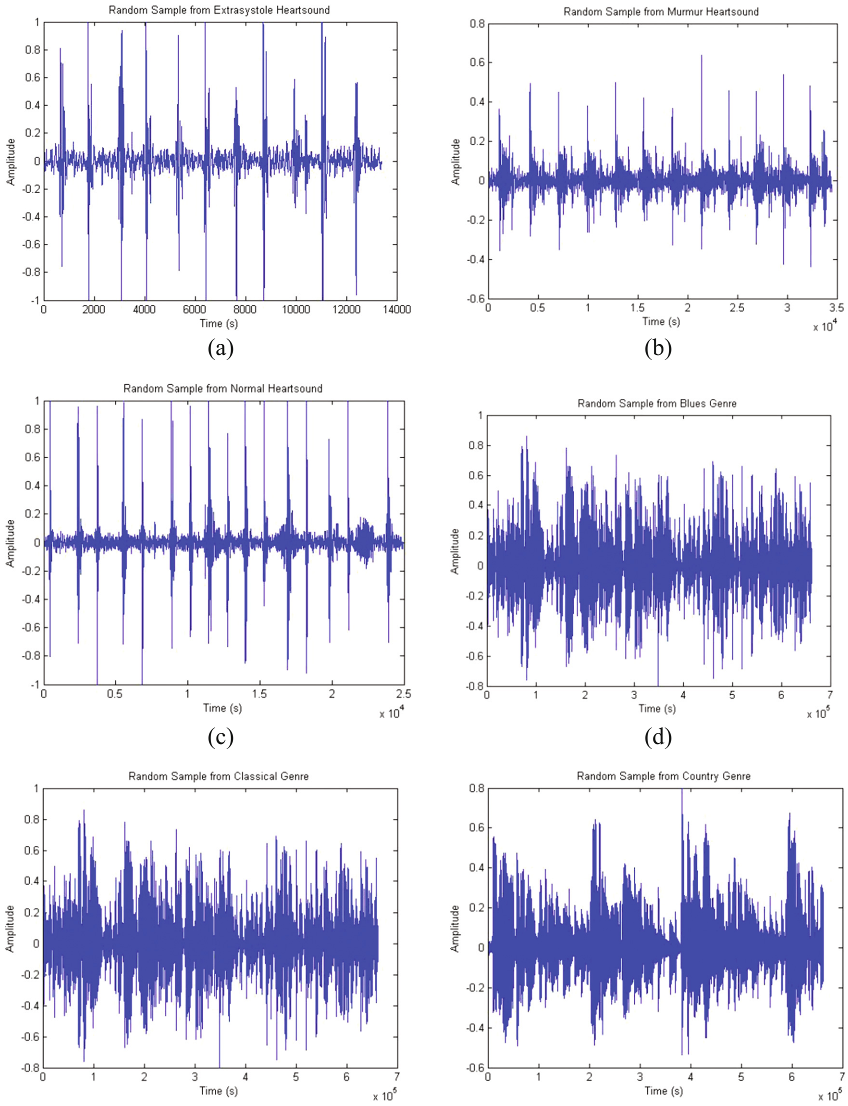
S.K. Kopparapu and K.K. Bhuvanagiri [13] proposed a modification in filter bank to correlate the MFCC features extracted at different sampling frequencies. The extracted features are then used to build acoustic models in the form of Hidden Markov Models. Zhai et al. [14] proposed a modified MFCC with sensitive frequency filter bank combined with HMM for classification of detected loose particles in sealed electronic devices. H. Lei and E.L. Gonzalo [15] studied the speaker discriminative power of mel, antimel, and linear cepstral coefficients in different speech regions such as nasal, vowel, and non-nasal consonant. The study shows LFCC having better relative EER improvement over MFCC in nasal and non-nasal consonant regions, and MFCC gave the relative improvement in vowel region. Loong, Justin Leo Cheang, et al. [16] demonstrated feature extraction in heart sounds using MFCC and LFCC and found MFCC system as superior to LFCC in recognition of heart sounds. X. Zhou et al. [17] compared the performance between MFCC and LFCC on a speaker recognition system and concluded that the LFCC outperformed on MFCC consistently due to its superior performance on female trails by improved capturing of the spectral characteristics in the high-frequency region. However, the result showed LFCC has only some advantage in reverberant speech and similar performance with MFCC in babble noise while poor performance in white noise. B.J. Shannon and K.K. Paliwal [18] studied the significance of spacing of filter banks using Mel scale, Bark scale, and Linear scale and concluded that the filter spacing did not provide any statistical improvements if training and testing conditions match. D. A. Reynolds [19] compared acoustic features in speaker identification task and found no significant difference between MFCC and LFCC. In literature [20], LFCC is more robust in whisper speech identification. Lawson et al. [21], studied acoustic features with the help of a classifier named GMM-super vectors and found LFCC has superior accuracy over MFCC in the majority of cases.

Although many modifications in filter banks, including linear filter bank, was proposed by many researchers, a detailed study on the statistical significance of linear filter bank is absent in the literature. Also, the findings in different research gave contradicting conclusions which motivated us to examine the statistical significance of MFCC and LFCC for various database and provide a conclusion. In this paper, mel-scale and linear scale filter bank was designed to analyse the interclass variability and intraclass similarity of the cepstral features extracted from different classes of Heart Sounds (HS) and Music Signals (MS) from two different databases. The statistical significance of the features is studied using Kruskal-Wallis test, and the various results are plotted and described in detail in the following sections.

2 Database

There are two different datasets used in the study of the statistical significance of Mel and Linear Cepstral coefficients. The first dataset [22], consists of 3 classes of heart sounds namely Extra systole, Murmur and Normal. There are 25 samples from each class considered for the analysis. The plots of the random specimen of each class of Heart Sounds are depicted in the Fig. 1(a), (b), and (c). Correspondingly, the second dataset is GTZAN dataset [23] which consists of 10 different classes of musical genres.

In the study, fifteen specimens from five different categories namely blues, classical, country, disco and hip-hop were selected for analysis. The Fig. 1(d), (e) and (f) illustrates the plots of randomly selected Music sample different classes. By analyzing the waveforms, the difference in the pattern of heart sounds and Musical signals are quite evident. Also, it is obvious to observe that waveforms from different classes of the same database display dissimilarities in the pattern.



(a) (b) (c) (d) (e) (f)

Fig. 1. (a) Extrasystole Heart Signal. (b) Murmur Heart Signal. (c) Normal Heart Signal. (d) Blues Music Signal. (e) Classical Music Signal. (f) Country Music Signal

3 Methodology

3.1 Methodology for Computing Cepstral Coefficients

The specimen needs to be transformed into the frequency domain to compute the cepstral coefficients. But before computing the spectrum, the audio specimens are preconditioned with normalization. Since the technical specification of the sensors used for the acquisition of audio signal and the level of amplification employed in each database remain unknown, to standardize the amplitudes of the heart sounds, the samples normalized to a range between −1 and +1 as in (1) (Fig. 2).

$$x_n(n) = \frac{x(n)}{max|x(n)|} \tag{1}$$

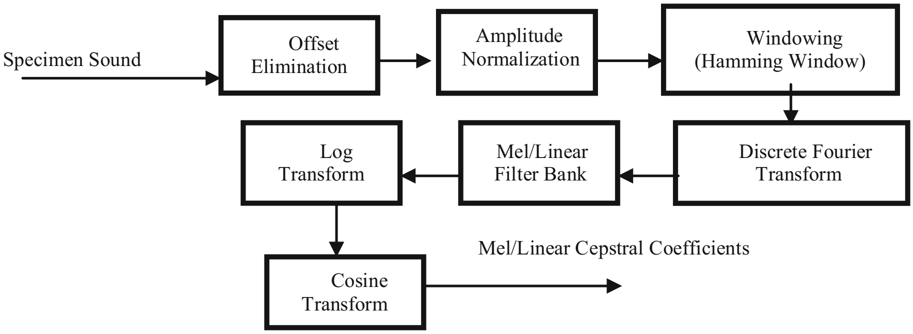


Fig. 2. Block diagram of methodology

Given the x(n) is the heart sound specimen, sampled at a rate $1/f_s$ and comprising N discrete samples, $1 \leq n \leq N$. The spectrum of the normalized specimen is computed using the FFT algorithm. Even though FFT allows fast computation of the spectrum, in FFT the technical issues induced by inadequate spectral resolution become too apparent, different from the direct computation of Discrete Fourier Transform (DFT). The zero padding improves the spectral resolution by interpolating intermediate bins in the frequency vector. To maintain adequate spectral resolution the number of samples in the specimen is increased to at least two times of its sampling rate 'f_s', through zero padding. The sound samples are windowed with Hamming window [24] before the computation of FFT to suppress the ripples which could be induced in the spectrum during the computation of FFT, because of the zero padding. The normalized sound signal after windowing is in (2).

$$\mathrm{x_w(n) = x_n(n)w(n)} \tag{2}$$

where the Hamming window is given by

$$W(n) = \begin{cases} 0.54 - 0.46\cos(2\pi n/N - 1) & \text{for} \quad 0 \leq n \leq N \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

$X_w(n)$ is the signal after normalization, zero-padding and windowing, obtained from the raw specimen signal. The spectrum of the windowed signal is computed as,

$$x(k) = \sum_{n=1}^{N} x_w(n) e^{\frac{-j2\pi kn}{N}}, \quad 1 \leq k \leq N \text{ and } 1 \leq n \leq N \tag{4}$$

We describe a filter bank having M filters (m = 1 to M), with m is triangular filter, given

$$\begin{array}{ll} H[m,k] = 0 & \text{if } k \ < f[m-1] \\ H[m,k] = (k - f[m-1])/(f\ [m] - f[m-1]) & \text{if } f[m-1] \leq k \ \leq f[m] \\ H[m,k] = (f\ [m+1] - k)/(f\ [m+1] - f[m]) & \text{if } f[m] \leq k \leq f[m+1] \\ H[m,k] = 0 & \text{if } k \ > f[m+1] \end{array} \tag{3}$$

satisfying

$$\sum_{m=1}^{M} H[m,k] = 1,\ k = 0, 1, \ldots, N - 1.$$

Let and f_h and f_l be the highest and lowest filter bank frequencies in Hz set by analysing the frequency response of the sensor and range of the audio samples. Let F_s be the sampling rate of the specimen sound in Hz. In mel-scale, the boundary points f [m] are evenly spread out:

$$f[m] = \left(\frac{N}{F_s}\right) B^{-1}\left(B(f_l) + m\frac{B(f_h) - B(f_l)}{M + 1}\right) \tag{8}$$

where,

$$B(f) = 1125\ln\left(1 + \frac{f}{700}\right) \tag{9}$$

And

$$B^{-1}(b) = 700\left(e^{(b/1125)} - 1\right) \tag{10}$$

Log-energy is calculated by,

$$S[m] = \ln\left\{\sum_{k=0}^{N-1} |X[k]|^2 H[m,k]\right\}, \quad 0 < m \leq M \tag{11}$$

Discrete cosine transform of the M filter outputs are calculated to obtain the MFCC:

$$c(n) = \sum_{m=0}^{M-1} S[m]\cos(\pi n(m - 0.5)/M), \quad 0 \leq n < M \tag{12}$$

In the Linear Frequency filter bank, the bandwidth of the individual triangular filters remains constant throughout the frequency range. In contrary to Mel-Cepstrum, the slope of the response of the triangular window remains the same in both low and high frequency. Hence, the levels to which two closely spaced frequency components mapped by the triangular filter response at higher frequencies and the levels to which two closely spaced frequency components mapped at lower frequencies remains the same.

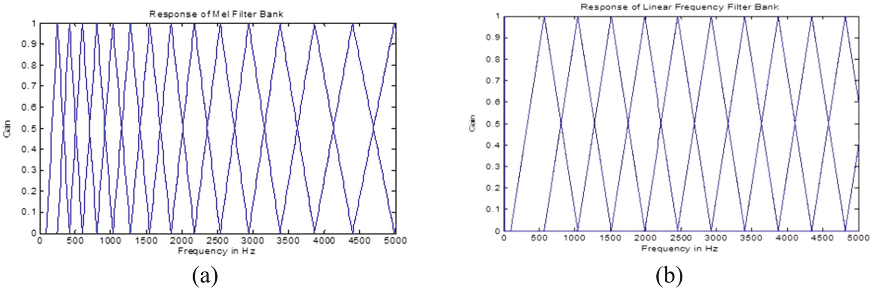


Fig. 3. (a) Response of Mel-Filter Bank. (b) Response of Linear Filter Bank

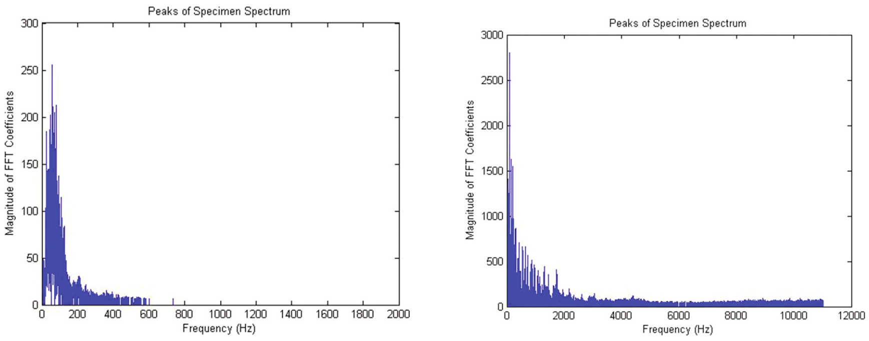
The Mel filter bank, whose response is as shown in Fig. 3a, is replaced with the linear frequency filter bank. Figure 3b shows the response of linear filter bank. The frontier points f[m] are homogenously and linearly spread out in the linear-scale:

$$f[m] = \left(\frac{N}{F_s}\right)\left(f_l + m\frac{f_h - f_l}{M + 1}\right) \tag{13}$$

4 Results and Discussion

4.1 Methodology for Computing Cepstral Coefficients

As mentioned in Sect. 3.1, the test is conducted to find the range of different specimen signals to attain the value of highest filter bank frequency. Results of the maximum



(a) (b)

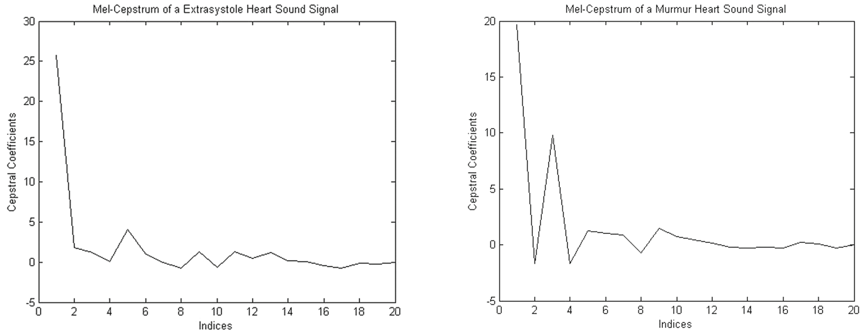
Fig. 4. (a) Peaks of Spectrum of Hear Sounds (b) Peaks of Spectrum of Music Signal

frequency test on a random sample from both the database are as shown in Fig. 4. It can easily be observed that majority of the signal coefficients lie below 1200 Hz in the case of heart sound signals and 12000 Hz for Music signals. Based on the test done in MATLAB the fl and fh was fixed as 50 and 1200 Hz for heart sound signals and 50 and 12000 Hz for Music genre specimens.

50 Hz lower frequency was taken due to the possibility of very low-frequency interferences that could occur during the recording process.

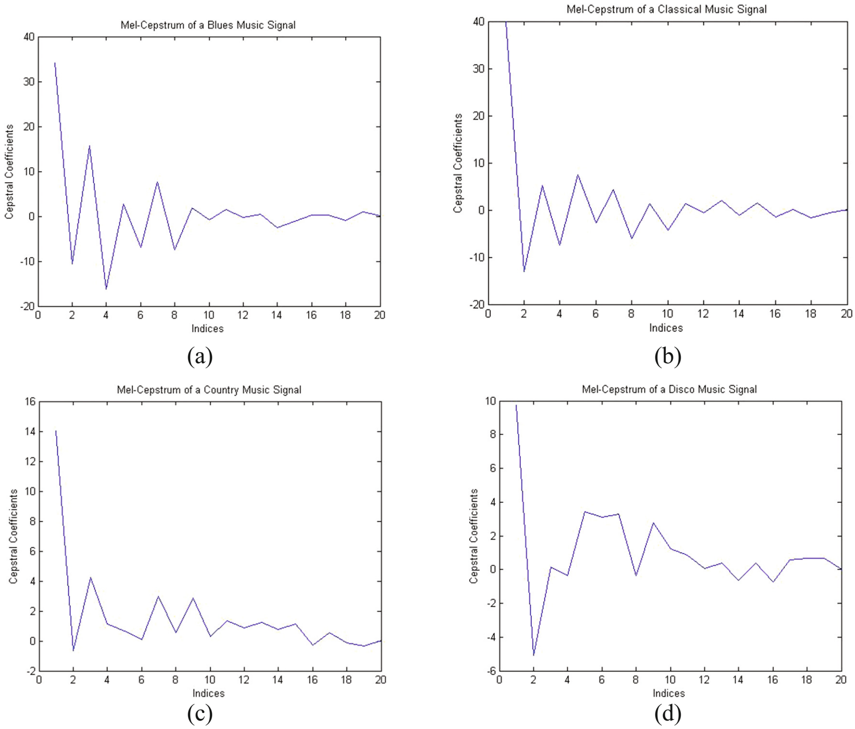
Mel-Cepstrum of heart sounds signals of different classes, such as Extrasystole, Murmur and Normal, plotted are analysed. The Mel-Cepstrum of randomly selected sample from the different database is as shown in Fig. 5. The inter-class variations in Mel-Cepstral Coefficient values of different classes in the Heart sound database are evident from qualitative analysis of the Figs. 5a and b. The intra-class similarity of Mel-Cepstrum was also observed from the analysis of results.

Similarly, Mel Cepstrum of randomly selected specimen from music genre database of different classes such as Blues, Classical, Country and Disco are as shown in Fig. 6.



(a) (b)

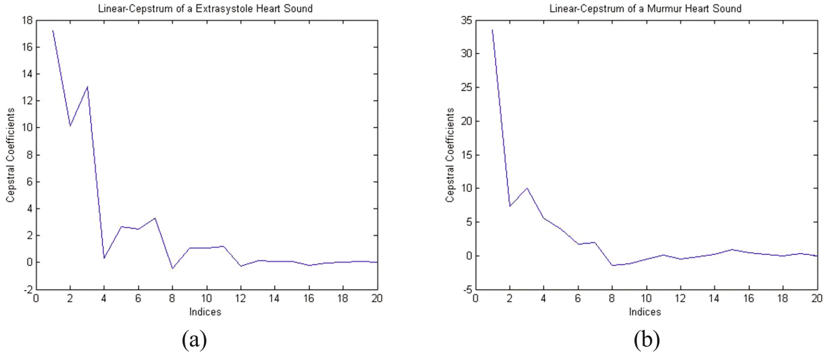
Fig. 5. (a) Mel-Cepstrum of Extrasystole HS. (b) Mel-Cepstrum of Murmur HS



(a) (b) (c) (d)

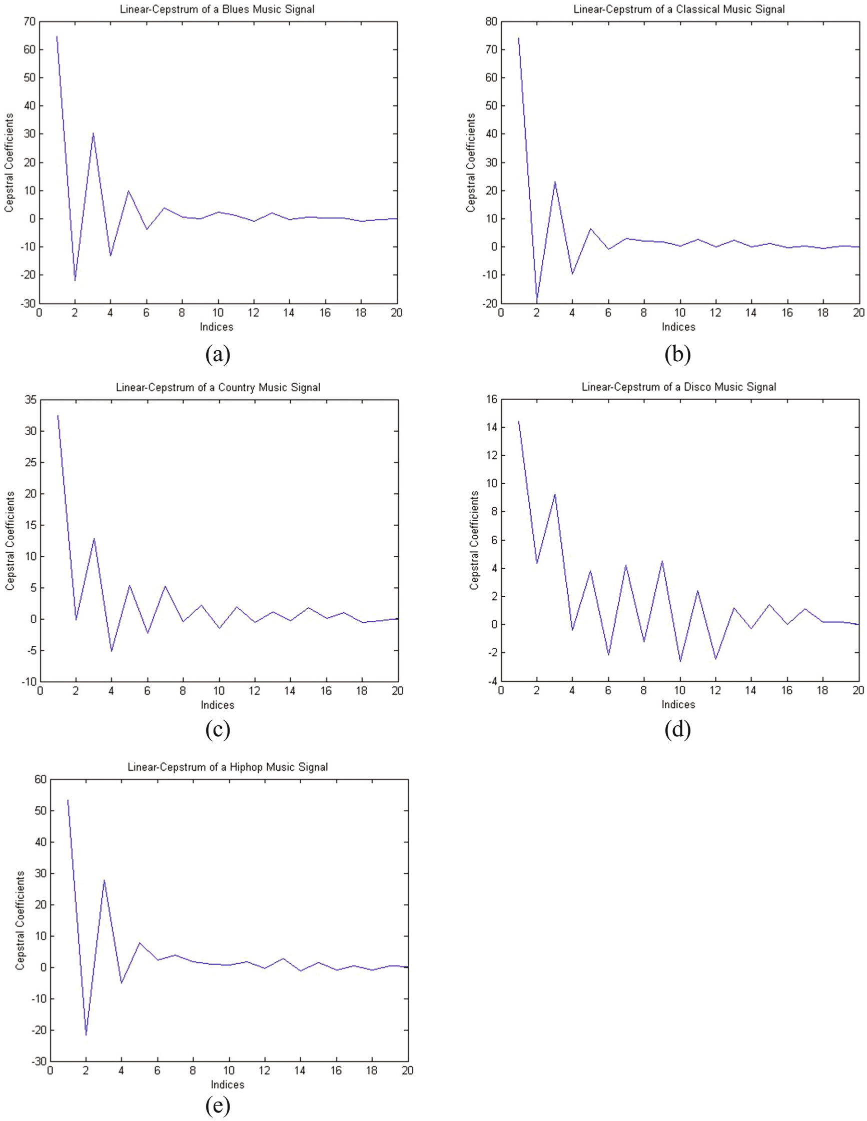
Fig. 6. (a) Mel-Cepstrum of Blues MS. (b) Mel-Cepstrum of Classical MS. (c) Mel-Cepstrum of Country MS. (d) Mel-Cepstrum of Disco MS

The procedure repeated with Linear Cepstrum for both the databases were analysed. The Linear Cepstrum of randomly selected Heart Sound signals and Music Signals from the corresponding dataset are given in Figs. 7 and 8 respectively



(a) (b)

Fig. 7. (a) Linear-Cepstrum of Extrasystole HS (b) Linear-Cepstrum of Murmur HS


Linear-Cepstrum of a Blues Music Signal

(a)

Linear-Cepstrum of a Classical Music Signal

(b)

Linear-Cepstrum of a Country Music Signal

(c)

Linear-Cepstrum of a Disco Music Signal

(d)

Linear-Cepstrum of a Hiphop Music Signal

(e)

Fig. 8. (a) Linear-Cepstrum of Blues MS. (b) Linear-Cepstrum of Classical MS. (c) Linear-Cepstrum of Country MS. (d) Linear-Cepstrum of Disco MS. (e) Linear-Cepstrum of Hip-hop MS

Interclass variability and intraclass similarity of both Mel and Linear Cepstrum can be qualitatively analysed by studying the plots of Cepstrum, which is as shown in Figs. 5, 6, 7 and 8. Qualitative analysis revealed that both MFC and LFC exhibit similarity among within the class specimen and variability among between the class sound specimens. Also, as it can be inferred from the cepstral coefficient plots (Figs. 5, 6, 7 and 8), instead of twenty Cepstral coefficients; only 13 or 14 Cepstral coefficients are significant. Thus the feature dimensionality can be reduced. In the study, only 13 coefficients are considered since higher coefficients didn't show significant variations and remained almost constant.

4.2 Analysis of Kruskal-Wallis Test on Heart Sounds Database

Please try to avoid rasterized images for line-art diagrams and schemas

It can be analysed from Table 1 that, the Mel-coefficients 1, 2, 4, 5, 8, 12 and 13 offer lower p-values than Linear-coefficients, which implies that they may provide better statistical significance.

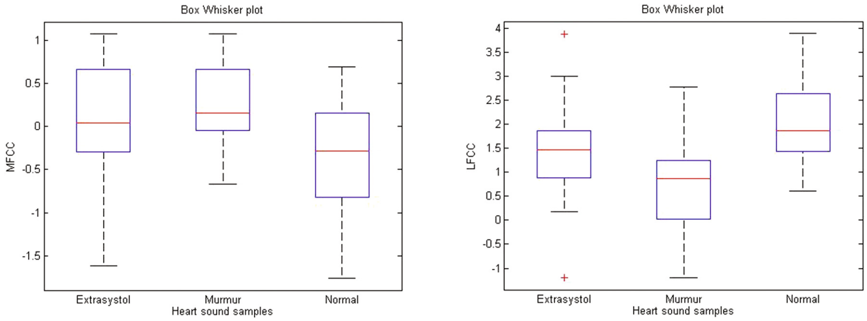
Table 1. p-values from Kruskal Wallis test of heartsounds database

Index of the coefficients	p-values of Mel-Cepstral analysis	p-values of Linear Cepstral analysis
1	0.1797	0.5152
2	0.0297	0.0601
3	0.4434	0.214
4	0.2157	0.4843
5	0.0128	0.12
6	0.1451	0.1397
7	0.0875	0.0143
8	0.0298	0.3843
9	0.879	0.0001
10	0.973	0.3962
11	0.77	0.4712
12	0.0056	0.0136
13	0.165	0.6754

The two lowest p-value obtained in the test using MFCC were 0.0056 and 0.0128 while that of LFCC are 0.0001 and 0.0136. Whereas, the maximum values are 0.973 and 0.879 for MFCC and 0.6754 and 0.5152 for LFCC. From the Table 1, in the case of heart sounds, MFCC gives the best and worst statistically significant feature. Moreover, the mean of p-values of all Mel-coefficients are found to be 0.30279, and that of Linear-coefficients are 0.2683385.

Now, on analysing box-plots, none of the MFCCs of the signals in the database offer an effective separability between any groups. The twelfth MFCC has the highest statistical significance in the Mel scale group, and the ninth LFCC has the highest statistical significance in the linear scale group. The box plots of most statistical significant coefficients in MFCC and LFCC are as shown in Fig. 9.

However, from the plot, it is evident that the most significant LFCC provides better separability than the most significant MFCC.



(a) (b)

Fig. 9. (a) Box Plot of MFCC. (b) Box Plot of LFCC

4.3 Analysis of Kruskal-Wallis Test on Music Genre Database

Please try to avoid rasterized images for line-art diagrams and schemas. Whenever possible, use vector graphics.

It can be observed from Table 2 that, Mel-coefficients 1, 3, 4, 5, 6, 7, 8, 10, 11, 12 and 13 offer lower p-values than Linear-coefficients, which implies that they may provide considerable statistical significance.

The two lowest p-value obtained in the test using MFCC were 1.71231e-11 and 4.83909e-11 while that of LFCC are 3.67083e-11 and 8.54791e-10. Whereas, the maximum values are 0.0266 and 0.0008 for MFCC and 0.2644and 0.1591 for LFCC. From the table, in the case of Music audio signal, MFCC gives the best statistically significant features for classification. In this case, the LFCC contributes the least significant feature providing the maximum p-value. Moreover, the mean of p-values of all

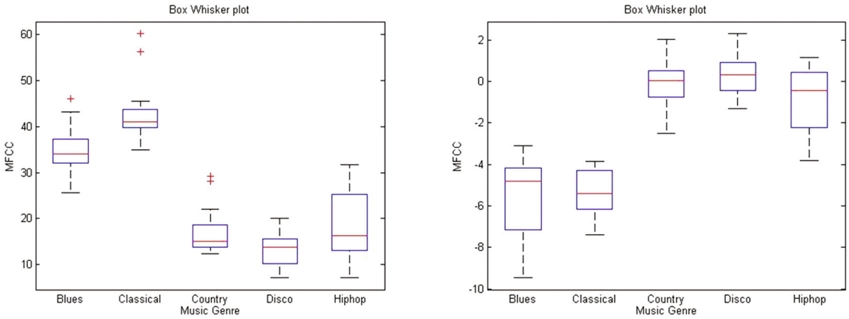
Table 2. P-values of music genre database

Index of the coefficients	p-value from Mel-Cepstral analysis	p-value from Linear Cepstral analysis
1	1.71231e-11	3.67083e-11
2	1.57659e-7	8.96314e-10
3	9.61781e-9	1.6195e-8
4	4.1682e-9	0.0006
5	3.13264e-6	0.0674
6	5.22077e-9	1.42572e-5
7	4.23137e-6	9.65585e-6
8	4.83909e-11	8.54791e-10
9	0.0266	3.66986e-6
10	1.01838e-7	0.1591
11	0.0008	0.0005
12	1.13918e-5	0.0107
13	0.0001	0.2644

Mel-coefficients are found to be 0.002116849, and that of Linear-coefficients is 0.038671354.

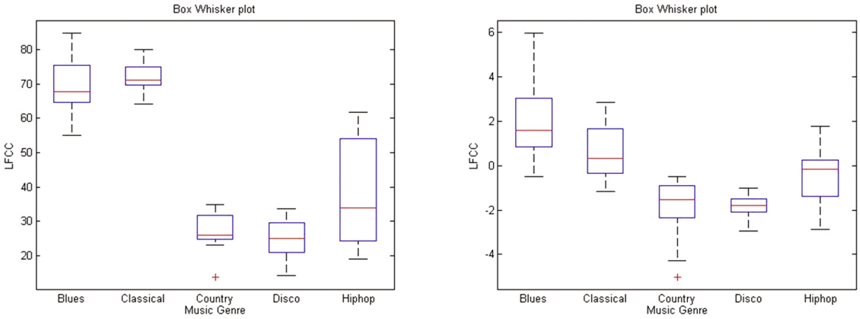
Now, on analysing box-plots, we can observe that many of the MFCCs and LFCCs of the acoustic signals in Music Genre database offer an effective separability between the groups. The first and eighth features have highest statistical significance in the Mel scale group the linear scale group. The box plots of most statistical significant coefficients of MFCC and LFCC are given in Figs. 10 and 11 respectively.

In the box plot analysis, it is visible that both the MFCCs and LFCCs can provide good statistically significant features for classification. However, on the closer analysis, MFCC features have an upper hand in case of separability between the classes.



(a) (b)

Fig. 10. Box Plots of MFCC



(a) (b)

Fig. 11. Box Plots of LFCC

5 Conclusion

In this paper, the statistical significance of Mel frequency cepstral coefficient (MFCC) and Linear frequency cepstral coefficients (LFCC) has been investigated with the help of two databases from different modalities. First, MFCC and LFCC features are extracted from both the databases and are plotted for qualitative investigation. Later, Kruskal-Wallis test was performed on both the feature sets for the detailed inference. The studies have demonstrated both MFCC and LFCC proved to be better than the other in two different databases. It is quite evident from results, neither MFCC nor LFCC can be stated as superior or inferior based on testing on a single dataset. The statistical significance provided by the features in case of MFCCs and LFCCs depend in large part on the variation of different attributes of the signal. Hence, careful study about the signal and the dependency of various attributes of the filter bank structure must be conducted before finally fixing the feature for classification. In future, an automated adaptive classifier, which automatically selects the feature by identifying the characteristics of the signals, must be developed. Besides, there are vast opportunities for researchers to implement automated classification algorithms dedicated for the particular type of signals by studying the signal attributes and selection of feature set, which is best appropriate for the signals.

Conflict of Interest: The authors declare that there is no conflict of interest regarding the publication of this paper.

References

1. Davis, S.B., Mermelstein, P.: Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. IEEE Trans. Acoust. Speech Sig. Proc. **28**, 357–366 (1980)
2. Mermelstein, P.: Distance measures for speech recognition–psychological and instrumental. In: Joint Workshop on Pattern Recognition and Artificial Intelligence (1976)
3. O'Shaughnessy, D.: Invited paper: Automatic speech recognition: history, methods and challenges. Pattern Recogn. **41**(10), 2965–2979 (2008). http://dx.doi.org/10.1016/j.patcog.2008.05.008, ISSN 0031-3203
4. Jo, J., Yoo, H., Park, I.C.: Energy-efficient floating-point MFCC extraction architecture for speech recognition systems. IEEE Trans. Very Large Scale Integr. VLSI Syst. **24**(2), 754–758 (2016). doi:10.1109/TVLSI.2015.2413454
5. Sahidullah, M., Tomi, K.: Local spectral variability features for speaker verification. Digital Signal Process. **50**, 1–11 (2016). http://dx.doi.org/10.1016/j.dsp.2015.10.011, ISSN 1051-2004
6. Anzar, S.M., Amala, K., Remya, R., Ashwin, M., Ajeesh, P.S., Mohammed, S.K., Febin, A.: Efficient online and offline template update mechanisms for speaker recognition. Comput. Electr. Eng. **50**, 10–25 (2016). http://dx.doi.org/10.1016/j.compeleceng.2015.12.003, ISSN 0045-7906
7. Nalini, N.J., Palanivel, S.: Music emotion recognition: the combined evidence of MFCC and residual phase. Egypt. Inform. J. **17**(1), 1–10 (2016). http://dx.doi.org/10.1016/j.eij.2015.05.004, ISSN 1110-8665
8. Sharma, A., Kaul, S.: Two-stage supervised learning-based method to detect screams and cries in urban environments. IEEE/ACM Trans. Audio, Speech, Lang. Proc. **24**(2), 290–299 (2016). doi:10.1109/TASLP.2015.2506264
9. Shan, Y., Zhang, Z., Yang, P., Huang, K.: Adaptive slice representation for human action classification. IEEE Trans. Circuits Syst. Video Technol. **25**(10), 1624–1636 (2015)

10. Nandini, S., Sahidullah, M.d., Goutam, S.: Lung sound classification using cepstral-based statistical features. Comput. Biol. Med. (2016). http://dx.doi.org/10.1016/j.compbiomed.2016.05.013, ISSN 0010-4825
11. Sunita, C., Ping, W., Chu, S.L., Anantharaman, V.: A computer-aided MFCC-based HMM system for automatic auscultation. Comput. Biol. Med. **38**(2), 221–233 (2008). http://dx.doi.org/10.1016/j.compbiomed.2007.10.006, ISSN 0010-4825
12. Chen, T.E., Yang, S.I., Ho, L.T., Tsai, K.H., Chen, Y.H., Chang, Y.F., .Wu, C.C.: S1 and S2 heart sound recognition using deep neural networks. IEEE Trans. Biomed. Eng. **PP**(99), 1 (2017)
13. Sunil, K., Kopparapu, K., Kumar, B.: Recognition of subsampled speech using a modified Mel filter bank. Comput. Electr. Eng. **39**(2), 655–662 (2013). http://dx.doi.org/10.1016/j.compeleceng.2012.10.002, ISSN 0045-7906
14. Guofu, Z., Jinbao, C., Chao, L., Guotao, W.: Pattern recognition approach to identify loose particle material based on modified MFCC and HMMs. Neurocomputing **155**(1), 135–145 (2015). http://dx.doi.org/10.1016/j.neucom.2014.12.039, ISSN 0925-2312
15. Lei, H., Gonzalo, E.L.: Mel, linear, and antimel frequency cepstral coefficients in broad phonetic regions for telephone speaker recognition. In: INTERSPEECH, pp. 2323–2326, September 2009
16. Loong, J.L., Subari, K.S., Abdullah, M.K., Ahmad, N.N.: Comparison of MFCC and cepstral coefficients as a feature set for PCG biometric systems. World Acad. Sci. Eng. Technol. Int. J. Med. Health Biomed. Bioeng. Pharm. Eng. **4**(8), 335–339 (2010)
17. Zhou, X., Garcia-Romero, D., Duraiswami, R., Espy-Wilson, C., Shamma, S: Linear versus mel frequency cepstral coefficients for speaker recognition. In: 2011 IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU), Waikoloa, HI, pp. 559–564 (2011)
18. Shannon, B.J., Kuldip K.P.: A comparative study of filter bank spacing for speech recognition. In: Microelectronic Engineering Research Conference, p. 41 (2003)
19. Reynolds, D.A.: Experimental evaluation of features for robust speaker identification. IEEE Trans. Speech Audio Proc. **2**(4), 639–643 (1994). doi:10.1109/89.326623
20. Fan, X., Hansen, J.H.: Speaker identification with whispered speech based on modified LFCC parameters and feature mapping. In: 2009 IEEE International Conference on Acoustics, Speech and Signal Processing, 19 April 2009, pp. 4553–4556. IEEE (2009)
21. Lawson, A., Vabishchevich, P., Huggins, M., Ardis, P., Battles, B., Stauffer, A.: Survey and evaluation of acoustic features for speaker recognition. In: 2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 5444–5447, 22 May 2011
22. Bentley, P., Nordehn, G., Coimbra, M., Mannor, S.: The pascal classifying heart sounds challenge (chsc2011) (2011). http://www.peterjbentley.com/heartchallenge/index.html.2011
23. Tzanetakis, George, Cook, Perry: Musical genre classification of audio signals. IEEE Trans. Speech Audio Proc. **10**(5), 293–302 (2002)
24. John, G., Proakis, D., Manolakis, G.: Digital Signal Processing: Principles, Algorithms and Applications, 3rd edn. Prentice Hall, Upper Saddle River (1995)

Particle Filtering Technique for Fast Fading Shadow Power Estimation in Wireless Communication

S. Jaiyant Gopal, J.P. Anita(✉), and P. Sudheesh

Department of Electronics and Communication Engineering,
Amrita School of Engineering, Coimbatore, Amrita Vishwa Vidyapeetham,
Amrita University, Coimbatore, India
jaiyantgopal6@gmail.com,
{jp_anita, p_sudheesh}@cb.amrita.edu

Abstract. There is a crucial importance of estimation of fading power in a mobile wireless communication system. This estimation is used for many handoff algorithms, power control, and adaptive transmission methods. This estimation of power loss can be used to reduce discrepancies and provide better wireless communication service to the user. Until now the window based weighted sample average estimator was used because of its simplicity. But it has its own disadvantages and hence use of Kalman filtering and adaptive Kalman filtering was proposed. Based on an autoregressive model of shadow fading power, particle filter algorithm is proposed in this paper in order to increase the efficiency of estimation and to obtain accurate results. The simulation and analysis presented in this paper have provided promising and supporting results.

Keywords: Particle filter · Power fading · Shadow power estimation · Rayleigh

1 Introduction

The performance of any wireless communication system depends largely on the wireless channel environment present. In the modern cities and towns there are many obstructions in the form of tall buildings that disturb the mobile signal. The wireless communication system consists of a base station (BS) and a/many mobile stations (MS). The obstructions shadow the signal that is transmitted from BS and results in power drop at the MS side. This power drop is mainly caused due to motion of MS. This motion leads to a phenomenon called fading of signal power. There are many factors influencing fading in any wireless channel such as: multipath propagation, speed of mobile, speed of surrounding objects, transmission bandwidth [1] etc. Fading of signal power is of two types: shadow fading and multipath fading. Multipath fading is caused when the signal reaches MS by two or more paths. These signals will have same amplitude but will be of different phases. These are differences due to Doppler effect along different signal paths and time dispersion caused by propagation delays. Another type of fading is shadow fading. This is caused due to obstacles in the path of propagation such as tall buildings, hills, mountains etc.

S.M. Thampi et al. (eds.), *Advances in Signal Processing and Intelligent Recognition Systems*, Advances in Intelligent Systems and Computing 678,
DOI 10.1007/978-3-319-67934-1_10

Multipath fading has high frequency components and is referred as fast fading as shown Fig. 1 [2]. Multipath can be constructive or destructive [3] whereas shadow fading results in large scale power loss. Shadow fading power loss estimation is vital for handoff decision and power control. A handoff refers to transferring a call or data session from one cellular network to another or from one channel to another. To provide uninterrupted service to user, an effective handoff decision making is vital [4].

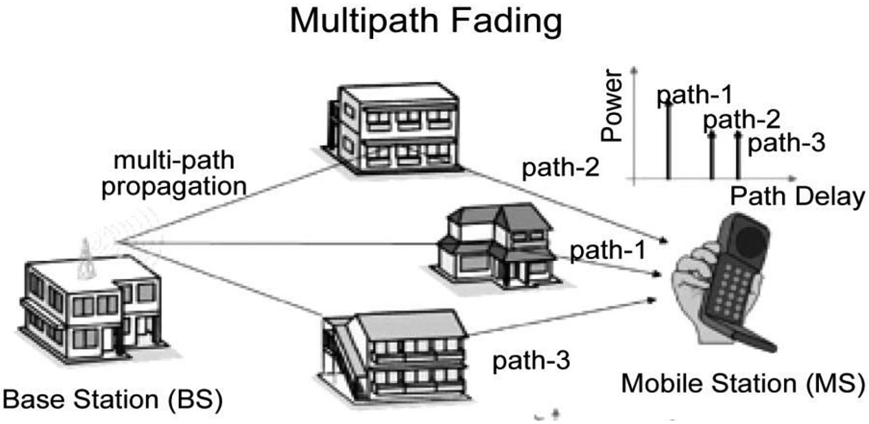


Fig. 1. An illustration for multipath fading loss due to different paths and buildings.

Accurate estimation of shadowing will enable better working of wireless communication system and compensate for signal degradation. Weighted sample average estimators, which is a window-based estimator of local mean power, is currently used in many commercial systems like GSM [4]. This estimator is utilized to filter multipath loss as Rayleigh noise [5]. But it assumes that the shadow power is constant. However, shadow power changes with time and this assumption leads to a lot of discrepancies. To estimate the shadow power loss in window based estimator the optimal window period should be known, which not only depends on MS velocity but also the sampling period and other shadow fading characteristics [6]. The lack of a consistent technique provides motivation to use methods with high fidelity. Here, the particle filtering (PF) based estimation method of local mean shadow power at mobile station is proposed over other traditional estimators such as window based estimator, Kalman filtering (KF) based estimator, adaptive Kalman filtering based estimator etc. The particle filtering algorithm based estimation of shadow fading power signal is based on the state space model (SSM) or hidden Markov model (HMM) [7], which is a statistical Markov model in which the system is said to be a Markov process (process where each state variable is independent) with unobserved (hidden) states [8]. In simple Markov models, all the states are directly visible and computable. In HMM there are hidden states but the output is dependent on the measurement state. Using these visible states prediction and estimation of hidden states can be done. By simulation results it is verified that PF based estimator is better than other methods in producing accurate results.

2 Signal Model and Problem Statement

The channel can be represented with the relationship of shadow fading power, received power and multipath fading power represented by $s(t)$, $l(t)$ and $w(t)$ respectively.

Equation (1) gives us the relation between received power, multipath fading and shadow fading.

$$l(t) = |w(t)|^2 s(t) \tag{1}$$

s(t) has to be estimated, so first the multipath fading power loss is modeled as a Rayleigh noise. Here it is dealt with time varying power signals and logarithm is used to represent signals

$$L(t) = 10log(l(t)) \quad S(t) = 10log(s(t) \qquad W(t){:} = 10log(w(t).$$

Thus Eq. (1) modified and shown as Eq. (2)

$$L(t) = S(t) + W(t) \tag{2}$$

The estimation is implemented in discrete time domain. So every time varying signal can be sampled with a sampling period of Ts.

$L(t){:} = L(nTs)$ also $S(t){:} = S(nTs)$. On the contrary, multipath fading power signal can be modeled as Rayleigh noise thus:

$$w(t) = \left(\frac{1}{\mathrm{R}}\right) \lim_{R\to\infty} \sum\nolimits_1^R br \, * \, e^{i(wD*Cos(\theta r)t + \varnothing r)} \tag{3}$$

Multipath loss modeled as Rayleigh Noise is shown in Eq. (3) where

- wD $= \frac{2\pi v}{\lambda}$ with v = velocity of Mobile Station
- R is the number of independent paths in multipath power, usually 20
- br are the gains corresponding to each R
- θr is the angle between the incoming waves and the mobile antenna
- $\varnothing r$ are phase random variables

By this model the multipath loss is obtained but the objective is to estimate shadow power fading as it essentially plays a very important role in wireless communication. So it is assumed that S(n) takes autoregressive (AR) model.

To perform particle filtering, there is a need of measurement equation and update equation. This can also be defined as the state space model [2]. The coefficients of the 'A' matrix and 'H' matrix are a_1 and a_2. These are shown in Eqs. (4), (5) and (6).

$$X_c = \frac{-D}{\ln(\varepsilon_D)} \tag{4}$$

$$a_1 = e^{\frac{-vT_s}{X_{c1}}} \tag{5}$$

$$a_2 = e^{\frac{-vT_s}{X_{c2}}} \tag{6}$$

$$x_k = \begin{bmatrix} x_{k1} \\ x_{k2} \end{bmatrix} \tag{7}$$

Equation (7) demonstrates the *xk* matrix in which *x1k* and *x2k* represent the state of which estimation has to be done.

From this the state space equations are given in Eqs. (8) and (9) as:

$$x(k+1) = Ax(k) + Bu(k) + W(k) \tag{8}$$

$$l(k) = Hx(k) + V(k) \tag{9}$$

In matrix form the Eqs. (8) and (9) can be shown as Eqs. (10) and (11)
Where Eq. (10) represents update step and (11) represents measurement step

$$\begin{bmatrix} x1k \\ x2k \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ a1 & a2 \end{bmatrix} \begin{bmatrix} x1(k-1) \\ x2(k-1) \end{bmatrix} + Wk \tag{10}$$

$$\begin{bmatrix} l1k \\ l2k \end{bmatrix} = [1\,0] \begin{bmatrix} x1k \\ x2k \end{bmatrix} + Vk \tag{11}$$

Where:

a1 is the first shadow power coefficient
a2 is the second shadow power coefficient
X_c is the correlation distance
ε_{D} is the correlation coefficient of shadow process between two points
D distance between the two points in meters
Wk is the process noise
Vk is the measurement noise

3 Methods for Estimation

Many methods are there for estimation but only Kalman filtering and particle filtering based methods are discussed in this paper.

3.1 Kalman Filter

Kalman filtering (KF) based estimator is extensively proposed in [1, 8, 9, 12] in this paper with the methodology of KF based estimator and the equations related to its algorithm are also discussed.

This Kalman filtering algorithm utilizes the same state space model as given in Eqs. (8) and (9). The equations used in KF are listed below:

Time update ("Predict") equations are:

(1) Project the state ahead:

$$\widehat{x}_{i+1}^{-} = A_i \widehat{x}_i + \mathrm{B} u_i \tag{12}$$

(2) Project the error covariance ahead:

$$P_{i+1}^{-} = A_i P_i A_i^T + Q_i \tag{13}$$

Measurement update ("Correct") equations are:

(1) Compute the Kalman gain:

$$K_i = P_i^- H_i^T \left(H_i P_i^- H_i^T + R_i\right)^{-1} \tag{14}$$

(2) Update estimate with measurement z_i:

$$\widehat{x}_i = \widehat{x}_i^- + K\left(z_i - H_i \widehat{x}_i^-\right) \tag{15}$$

(3) Update the error covariance:

$$P_i = (I - K_i H_i) P_i^- \tag{16}$$

These steps are repeated for the required number of iterations, the inputs for this cycle is $\widehat{x}_i^-$ and P_i^- of the previous step.

In KF, the MMSE is found from data corrupted with AWG noise in measurement and prediction steps.

Where:

P_i^- represents the priori estimate error covariance matrix.
P_i represents the posterior estimate error covariance matrix.
K_i is the Kalman gain or blending factor that minimizes the posteriori error covariance.
R_i is the measurement error covariance matrix.

The goal is to obtain the posterior estimate $\widehat{x}_i$ as a linear combination of an a priori estimate $\widehat{x}_i^-$ and the difference between an actual measurement z_i and a measurement prediction $H_i \widehat{x}_i^-$ as given in the Eq. (15).

The difference $z_i - H_i \widehat{x}_i^-$ in Eq. (15) is called the measurement innovation, or the residual. This shows the difference between $H_i \widehat{x}_i^-$ and the actual measurement z_i [13]. The results obtained through KF method are less accurate compared to PF method.

3.2 Particle Filter

Particle filtering is a sequential Monte Carlo method of recursive estimation of any Hidden Markov Model (HMM) where knowledge about the state is obtained from measurement states with additional noise present. Particle filters are based on probability distribution representation of states by a set of samples (particles) as shown in Fig. 2 [8]. It has an edge over other methods as nonlinear systems can also be represented as set of particles, and multi-modal non-Gaussian density states [10]. This particle filtering algorithm is an efficient alternative to the Markov Chain Monte Carlo (MCMC) algorithms. This can be used to create Markov chains [11].

For predication and estimation of the posterior density function we have two models: *system model* and *measurement model* in the probabilistic form. In general, in a Bayesian approach we consider all models and state variations in probabilistic form [15]. The particle filter is a recursive filtering approach, has two stages namely prediction and update and both stages utilize the system model and measurement model respectively.

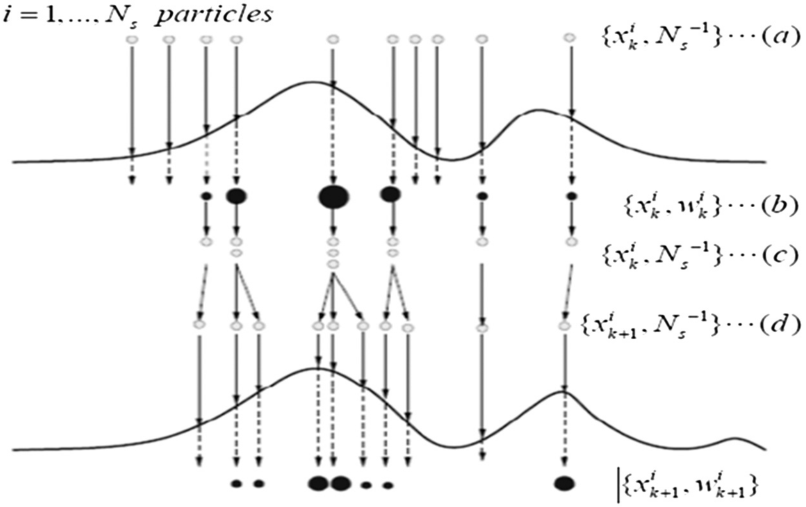

$i = 1, \ldots, N_s$ particles
$\{x_k^i, N_s^{-1}\} \cdots (a)$
$\{x_k^i, w_k^i\} \cdots (b)$
$\{x_k^i, N_s^{-1}\} \cdots (c)$
$\{x_{k+1}^i, N_s^{-1}\} \cdots (d)$
$\{x_{k+1}^i, w_{k+1}^i\}$

Fig. 2. Graphical representation of sequential importance sampling in particle filter algorithm.

For estimation, we define a vector X_k that represents the state of system at time

$$x_k = f_k(x_{k-1}, V_{k-1})$$

Here, V_k represents the Gaussian noise present.
The state vector x_k is defined by a nonlinear and time varying function f_k.

We can estimate the state variable using noisy measurements of z_k, which is governed by the equation

$$z_k = h_k(x_k, n_k)$$

Similar to the Kalman filter algorithm, here also we use the measurement states, which we denote by $z_{1:k}$, to find x_k. This is done by computing the probability distribution of $p(x_k|z_{1:k})$ which is done recursively in two steps

Prediction step:
$p(x_k|z_{1:k-1})$ is computed from $p(x_{k-1}|z_{1:k-1})$ at k – 1 instant by the Eq. (17)

$$p(x_k|z_{1:k-1}) = \int p(x_k|x_{k-1})p(x_{k-1}|z_{1:k-1})dx_{k-1} \tag{17}$$

Update step:
The prior estimate is updated with new measurements Z_k thus obtaining the posterior estimate state represented by Eq. (18)

$$p(x_k|z_{1:k}) \approx p(z_k|x_k)p(x_k|z_{1:k-1}) \tag{18}$$

But the problem is that we cannot directly compute or operate on these functions f_k and h_k thus we resort to approximate method which is sequential importance sampling (SIS) [14]. The goal of SIS is to estimate the posterior distribution at k-1 instant, $p(x_{0:k-1}|z_{1:k-1})$ with a set of samples called as particles and repetitively update these particles to obtain an estimation to the posterior distribution of the next step.

Particles are generated by taking samples from the priori distribution $q(x)$ and updating them according to the posterior distribution $p(x)$ [16]. Weight of each particle is represented by w_i which is obtained by the relation:

$$wi = \pi(xi)/q(xi)$$

Where $\pi(x)$ is a distribution proportional to $p(x)$

Thus importance sampling can be shown in Eq. (19) as,

$$p(x_{o:k-1}|z_{i:k-1}) \approx \sum\nolimits_{i=1}^{N} \omega_{k-1}^{i} \delta_{x_{0:k-1}}^{i} \tag{19}$$

Where δ_x is delta function centered at $x_{0:k-1}^{i}$

The weights of the particles are recursively updated using Eqs. (8) and (9); this is shown in Eq. (20),

$$\omega_k^i = \omega_{k-1}^i \left(\frac{p(z_k|x_k^i)p(x_k^i|x_{k-1}^i)}{q(x_k^i|x_{0:k-1}^i, z_{1:k})}\right) \tag{20}$$

In practice, we face the degeneracy problem [13] where the smaller weights give rise to errors. So we take the resampling process and check if the result is greater than N_{eff}.

$$N_{eff} = \frac{1}{\sum_{i=1}^{N}(\omega_k^i)^2}$$

Steps of particle filtering:

Step (1).
Initiate a set of $\mathrm{N_p}$ particles by using any random distribution.
Assign each particle with initial weight of $1/\mathrm{N_p}$.
Step (2).
Obtain the Nonlinear/linear update and measurement equations for estimation.
Step (3).
Using these equations estimate the *k*th step x_k value.
Step (4).
Update the weight of particles using

$$\bar{\omega}_n^i = \omega_{n-1}^i \left(\frac{1}{\sqrt{2\pi\sigma}}\right) e^{-\frac{|yn-xn|^2}{2\sigma^2}}$$

Step (5).
Normalize each weight
Step (6).
Calculate the effective particle size, if the effective particle size is larger than threshold then proceed to Step 7 otherwise resample and initialize the weights again.

$$\mathrm{N_{eff}} = \frac{1}{\sum_{\mathrm{p}=1}^{\mathrm{Np}}(\upomega_{\mathrm{k}}^{\mathrm{i}})^2}$$

Step (7).
Now approximate the estimated state value by multiplying and adding each particle to its corresponding weights.

4 Simulation and Results

In this section, we have studied the simulation results obtained. All simulations are done using the MATLAB software. From Fig. 3 it is inferred that the particle filter algorithm produces accurate results with high fidelity. The error margin is very low compared to other estimation methods. It is elucidated by Fig. 4 where the MSE is high for KF compared to PF even when the SNR values are increased. From Fig. 5 it can be inferred that increasing the number of iterations provides lesser error therefore giving better results. Similarly, it is illustrated that in Fig. 6 increasing the number of particles also increases the accuracy and reduces error but increasing number of particles and

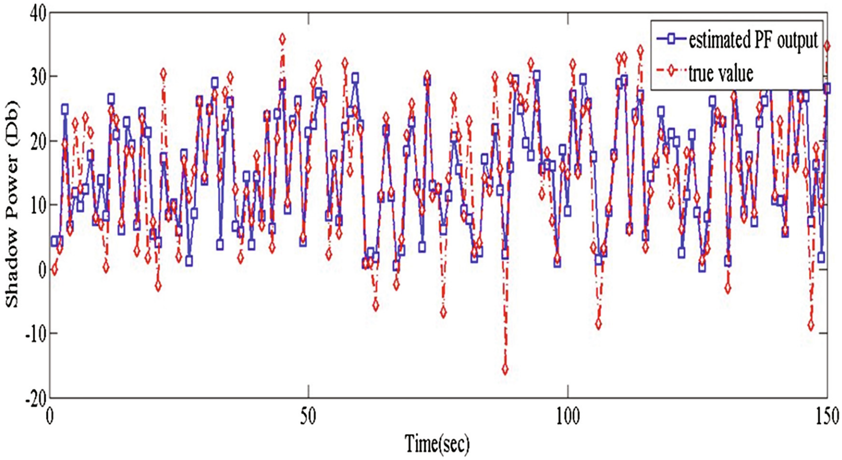


Fig. 3. Output result graph for estimated value and true value of shadow power in particle filter estimation method

Table 1. Computational time for constant number of iterations of particle filter algorithm in estimation of shadow power

No of iterations (T)	No of particles (N)	Computational time
100	100	15.232 s
100	150	26.116 s
100	200	37.284 s

Table 2. Computational time for constant number of particles of particle filter algorithm in estimation of shadow power

No of iterations (T)	No of particles (N)	Computational time
100	100	15.458 s
150	100	21.116 s
200	100	28.284 s

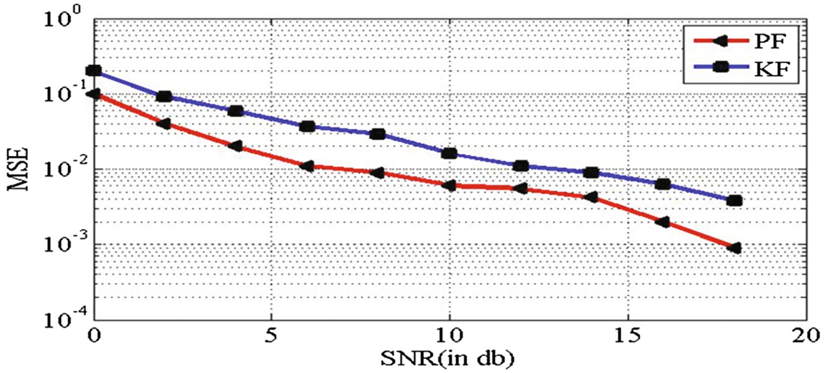


Fig. 4. Comparison of SNR versus MSE for KF method and PF method of shadow power estimation

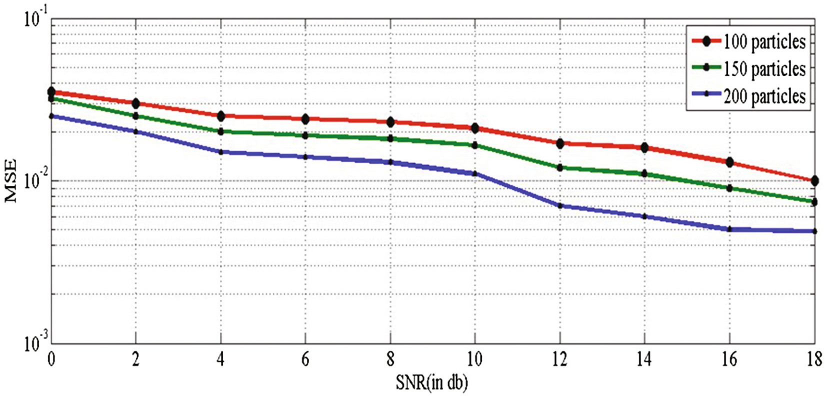


Fig. 5. Comparison of SNR versus MSE for different number of iterations of particle filter algorithm in estimation of shadow power

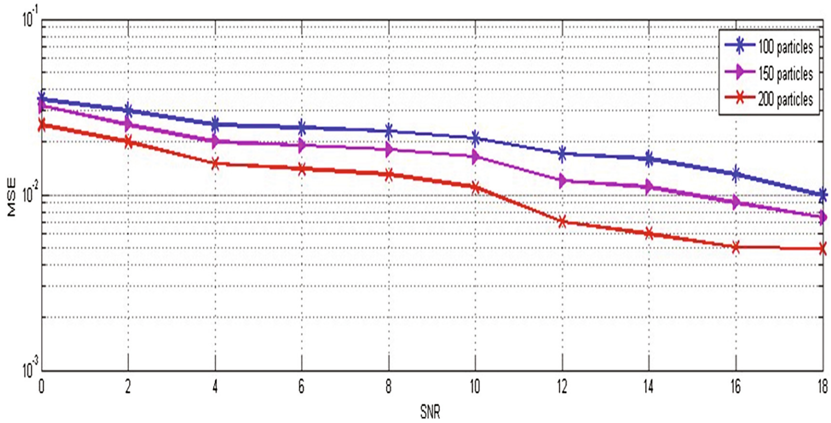


Fig. 6. Comparison of SNR versus MSE for different number of particles of particle filter algorithm in estimation of shadow power

number of iterations results in increase in computation time as illustrated in Tables 1 and 2. The increase in computational time as illustrated in Tables 1 and 2 gives us a tradeoff between accuracy, time taken for computation. If we want faster results there is a decline in accuracy of results, and for accurate results there is a raise in time taken.

5 Conclusion

The power estimation using particle filtering method has been discussed in this paper. The shadow power that is lost during transmission in a wireless channel mobile communication system is estimated in this paper. The different estimation techniques were discussed and the proposed particle filtering method has been proved to produce better and accurate results compared to other methods. This has been supported with simulation results using MATLAB. Thus, the conclusion of this study is that the PF method is superior compared to other estimation methods as there is high fidelity and statistical efficiency. The only disadvantage of this method is its high computational cost compared to other methods.

References

1. Jiang, T., Nicholas, S.D., Georgios-Giannakis, B.: Kalman filtering for power estimation in mobile communications. IEEE Trans. Wireless Commun. **2**(1), 151–161 (2006)
2. Kapetanovic, A., Mawari, R., Zohdy, M.: Second-order Kalman filtering application to fading channels supported by real data. J. Signal Inf. Proc. **7**(02), 61 (2016)
3. Rappaport, T.S.: Wireless Communications: Principles and Practice, vol. 2. Prentice Hall PTR, New Jersey (1996)
4. Grob, M.S., et al.: Method for robust handoff in wireless communication system. U.S. Patent No. 6,360,100, 19 March 2002
5. Sklar, B.: Rayleigh fading channels in mobile digital communication systems. I. Characterization. IEEE Commun. Mag. **35**(9), 136–146 (1997)
6. Salo, J., et al.: An additive model as a physical basis for shadow fading. IEEE Trans. Veh. Technol. **56**(1), 13–26 (2007)
7. Rabiner, L.: A tutorial on hidden Markov models and selected applications in speech recognition. In: Proceedings of IEEE (1989)
8. Kurt, T., et al.: Adaptive Kalman filtering for local mean power estimation in mobile communications. In: IEEE 64th Vehicular Technology Conference, VTC-2006 Fall. IEEE (2006)
9. Welch, G., Gary, B.: An Introduction to the Kalman Filter (1995)
10. Doucet, A., Nando, D., Freitas., Neil, G.: An introduction to sequential Monte Carlo methods. In: Sequential Monte Carlo Methods in Practice, pp. 3–14. Springer, New York (2001)
11. Arulampalam, S.M., et al.: A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. IEEE Trans. Signal Proc. **50**(2), 174–188 (2002)
12. Seshadri, V., Sudheesh, P., Jayakumar, M.: Tracking the variation of tidal stature using Kalman filter. In: International Conference on Circuit, Power and Computing Technologies (ICCPCT 2016) (2016)
13. Nair, N. Sudheesh, P. Jayakumar, M.: 2-D tracking of objects using Kalman filter. In: International Conference on circuit, Power and Computing Technologies (ICCPCT 2016) (2016)
14. Wei, J., et al.: A new particle filter object tracking algorithm based on dynamic transition model. In: IEEE International Conference on Information and Automation (ICIA). IEEE (2016)
15. Zhang, P.-L., et al.: Particle filtering based channel estimation in OFDM power line communication. J. China Univ. Posts Telecommun. **21**(5), 24–30 (2014)
16. Yang, T., Mehta, P.G., Meyn, S.P.: Feedback particle filter for a continuous-time Markov chain. IEEE Trans. Autom. Control **61**(2), 556–561 (2016)

A Novel Cyclic Convolution Based Regularization Method for Power-Line Interference Removal in ECG Signal

V.G. Sujadevi, K.P. Soman(✉), S. Sachin Kumar, and Neethu Mohan

Centre for Computational Engineering and Networking (CEN),
Amrita School of Engineering, Amrita Vishwa Vidyapeetham,
Amrita University, Coimbatore, India
sujapraba@gmail.com, kp_soman@amrita.edu

Abstract. Applying signal processing to bio-signal record such as electrocardiogram or ECG signals provide vital insights to the details in diagnosis. The diagnosis will be exact when the extracted information about the ECG is accurate. However, these records usually gets corrupted/contaminated with several artifacts and power-line interferences (PLI) thereby affects the quality of diagnosis. Power-line interferences occurs in the range close to 50 Hz/60 Hz. The challenge is to remove the interferences without altering the original characteristics of ECG signal. Since the ECG signals frequency range is close to PLI, several articles discuss PLI removal methods which are mathematically complex and computationally intense. The present paper proposes a novel PLI removal method that uses a simple optimization method involving a circular convolution based ℓ_2-norm regularization. The solution is obtained in closed form and hence computationally simple and fast. The effectiveness of the proposed method is evaluated using output signal-to-noise-ratio (SNR) measure, and is found to be state-of-the-art.

1 Introduction

The main objective of conducting electrocardiography is to get information about the function and structure of the heart in the form of electrocardiogram (ECG) signal. The ECG signal is more sensitive to different noises such as (1) high frequency noises (2) electrical interferences (3) motion artifacts (4) baseline wandering (5) muscle contractions etc. Out of which, power-line interference (PLI) noise is the most destructive noise related to ECG signal [1,2]. It is a high frequency noise and the ECG signal gets contaminated due to PLI during (1) recording (2) transmission (3) loops in the cables causes stray effect (4) improper grounding (5) electrical fluctuations etc. To infer correct diagnosis, the measured ECG signal reading must be free from noises. The presence of high PLI noise alters the quality of ECG signal which in turn results in the extraction of less

S.M. Thampi et al. (eds.), *Advances in Signal Processing and Intelligent Recognition Systems*, Advances in Intelligent Systems and Computing 678,
DOI 10.1007/978-3-319-67934-1_11

information. This will raise issue in finding the correct boundaries. The power-line interference signal is non-stationary in nature and it occur as a small band around the center frequency, 50 or 60 Hz and along its harmonics. Hence, it is essential to remove the PLI and it is a time consuming process. Numerous studies have been done elaborating the various methods used for PLI removal [3–10]. The most popular conventional method used for removing PLI is notch filtering [11]. The notch filter acts as the band-stop filter which removes the PLI at 50 Hz. However, during this filtering the information signal present at 50 Hz will also be removed and also introduces distortion. Due to this, notch filtering is not preferred even though its easy to implement. The use of adaptive filters removes the draw backs of notch filtering approach. However, the parameters used in the adaptive filtering method need to be changed according to the ECG signal. The article [12] provides a comparison among the adaptive and non-adaptive methods for PLI removal from ECG signal. Several other methods proposed for PLI removal are sliding DFT phase locking scheme [13], extended Kalman filter [14], least mean-square approach [15], modified VMD based method [16].

In this study, we used a circular convolution based regularization method for PLI removal. This is the first paper in this direction. Sinusoidal signal around 50 Hz (one need not know exact frequency of power signal) having same length as ECG signal is used for this purpose. The method involves only FFT computation of two sequences, an element-wise division of two vectors and an inverse FFT. Hence it is computationally fast. The output signal-to-noise-ratio (SNR) measure is found to be very high even for input signal with -17 dB as SNR. Further, it is observed that the output SNR does not vary much even if there is a big shift in power-line frequency. The paper is organized as follows: Sect. 2 discusses the proposed method used for removing the PLI from ECG signal. In Sect. 3, the experiments performed and the results obtained are discussed. The conclusion is provided in Sect. 4.

2 Proposed Approach - Matrix Free Filtering Method

The problem of PLI noise removal from an ECG signal can be expressed as,

$$y = x + p \tag{1}$$

where $y \in R^n$ is the noisy signal with PLI, $x \in R^n$ is the clean ECG signal, and $p \in R^n$ is the power-line interference signal. PLI signal is expressed as,

$$p(n) = a \cdot \cos(2\pi f(n\Delta t) + \phi) \tag{2}$$

where $f, a, \phi, \Delta t$ denotes the power line frequency, amplitude, phase, and sampling interval respectively. One of the state-of-the-art optimization method used for removing PLI is the following which performs an ordinary convolution with a derivative filter. The formulation is:

$$x^* = \arg\min_x \|y - x\|_2^2 + \lambda\|Dx\|_1 \tag{3}$$

Here, D is a matrix of size $(N-1) \times N$ and is defined as,

$$D = \begin{bmatrix} -1 & 1 & & & \\ & -1 & 1 & & \\ & & \cdot\ \cdot & & \\ & & & \cdot\ \ \cdot & \\ & & & -1 & 1 \end{bmatrix}_{(N-1)\times N} \tag{4}$$

The objective function consists of fidelity and regularization term. The fidelity term make sure that the denoised signal is close to the original signal. The second term performs regularization via a derivative matrix. This term penalizes high value of first-order slope at any point in the denoised signal. Using a control parameter λ, the trade-off between fidelity and regularization is controlled. In the proposed method sinusoidal signal of the frequency close to power-line signal frequency (50 Hz/60 Hz) is used. The length of sinusoidal signal must be same as original signal and chosen frequency need only to be close to power-line frequency. Even a deviation of 1 Hz will not cause deterioration of the performance. Convolution of two signals of same length with the use of DFT of the same length results in cyclic convolution [17]. Further, convolution in time domain is equivalent to element wise multiplication in frequency domain. This fact is used to model the optimization problem. The optimization problem is defined as ℓ_2-norm regularization,

$$X^* = \arg\min_X \|X - Y\|_2^2 + \lambda \|S. * X\|_2^2 \tag{5}$$

where X is the DFT of denoised signal x, Y is the DFT of noisy PLI signal y and S is the DFT of the sinusoidal wave with frequency around 50 Hz. Intuitively the formulation demands a denoised signal that is devoid of any 50 Hz or its shifted version. Formulation (5) can be simplified as,

$$X^* = \arg\min_X \sum_k (X_k - Y_k)^H (X_k - Y_k) + \lambda \sum_k (S_k X_k)^H (S_k X_k) \tag{6}$$

In the objective function, H denotes the Hermitian as the matrix is complex in nature after DFT operation. Now, the objective function is separable in k (elements of X) and hence we considers one separable term in k^{th} element as,

$$f(X_k) = (X_k - Y_k)^H (X_k - Y_k) + \lambda (S_k X_k)^H (S_k X_k) \tag{7}$$

λ is a regularization parameter. On differentiating the term with respect to X_k and putting equals to zero gives

$$\begin{aligned} &2(X_k - Y_k) + 2\lambda |S_k|^2 X_k = 0 \\ &X_k(1 + \lambda |S_k|^2) = Y_k \\ &X_k = \tfrac{Y_k}{(1+\lambda |S_k|^2)} \end{aligned} \tag{8}$$

The operation performed in Eq. (7) is element-wise division and holds an advantage against the matrix-based optimization methods and iterative methods. Here, the operation becomes a matrix-free operation. The denoised signal or PLI removed signal can be obtained using inverse Fourier transform operation. That is,

$$x = ifft(X) \tag{9}$$

The derivation uses following Lemma 1.

2.1 Lemma:1

If $f(Z) = Z\bar{Z}$, where $Z = Z_R + iZ_I$ and we are treating Z_R and Z_I as independent variables in the optimization, hence $f(Z)$ becomes a function of two variables. That is,

$$f(Z_R, Z_I) = (Z_R + iZ_I)(Z_R - iZ_I) = Z_R^2 + Z_I^2 \tag{10}$$

Now derivative of $f(Z_R, Z_I)$ can obtain as,
$\frac{\partial f}{\partial Z_R} = 2Z_R, \frac{\partial f}{\partial Z_I} = 2Z_I \Rightarrow \frac{\partial f}{\partial Z} = \frac{\partial f}{\partial Z_R} + i\frac{\partial f}{\partial Z_I} = 2(Z_R + iZ_I) = 2Z$

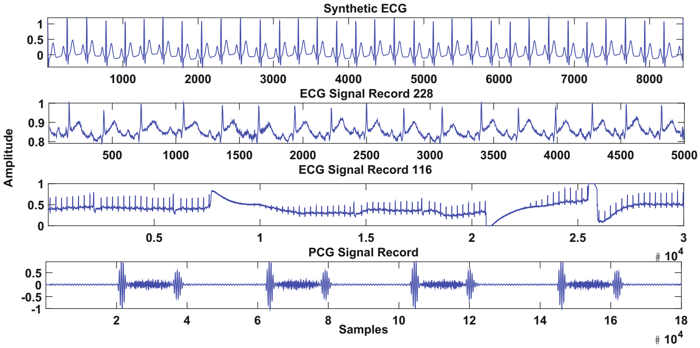


Fig. 1. Examples of signals considered for evaluations (a) Synthetic ECG signal [18], (b) ECG record 228 from [19], (c) ECG record 116 from [19], (d) PCG signal from [20].

3 Experiments and Results

The performance of the method proposed is evaluated on synthetic ECG signal [18], ECG records of MIT-BIH Arrhythmia database [19], and PCG signal [20]. Figure 1 shows examples of test signals used in this study. From the top, first row indicates the synthetic ECG signal, second and third signals corresponds to the ECG signal record 228 and 116 from MIT-BIH Arrythmia database. The other signal records used for the evaluation are record 108 and record 203. Instead of taking the entire signal for the evaluation, a small frame is considered. The fourth signal is a phonocardiogram (PCG) signal. Except the synthetic ECG signal, it can be observed that the rest of the signals already have noise at small level.

The performance of the proposed method is measured in terms of signal-to-noise ratio (SNR) metric. After denoising, better noise reduction is indicated by high SNR value.

3.1 Evaluations on Synthetic ECG Signals

The experiments are performed on synthetic ECG signal generated using synthetic ECG simulator, ECGSYN [18]. The sampling frequency is 256 Hz and a frame of 8500 samples is considered in this study. Figure 2 shows the performance on a syntectic ECG corrupted by 1 dB noise. As evident from the figure, the denoised signal has a high SNR of 37.8 dB and the error in estimation (difference between the clean signal and denoised signal) is negligible. To evaluate the approach, the synthetic sequence is corrupted with noise in various levels ranging from −15 dB to 28 dB. Table 1 shows the effectiveness of the proposed approach based on SNR matric. The regularization parameter, λ is fixed as 0.001 on all noise levels and it can be observed that the proposed approach obtained state-of-the-art output. Figure 3 depicts the denoising results of synthetic ECG at various noise levels.

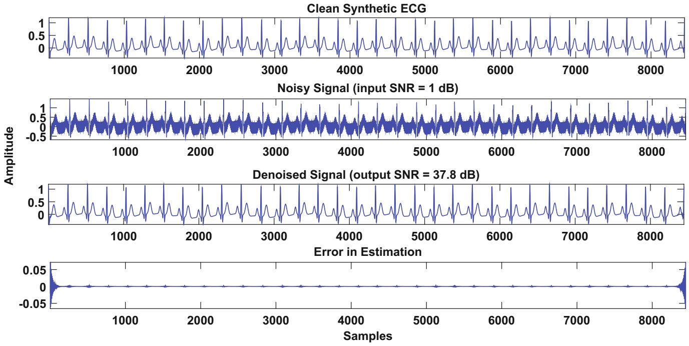


Fig. 2. Performance of the proposed approach on synthetic ECG corrupted by 1 dB noise (a) Clean ECG, (b) Noisy ECG, (c) Denoised ECG, (d) Error in estimation.

Table 1. SNR based evaluation of proposed method on synthetic ECG sequence ($\lambda = 0.001$)

Input SNR (dB)	Output SNR (dB)
28	58.95
12	52.03
−1	38.86
−5	34.80
−10	29.35
−15	24.91

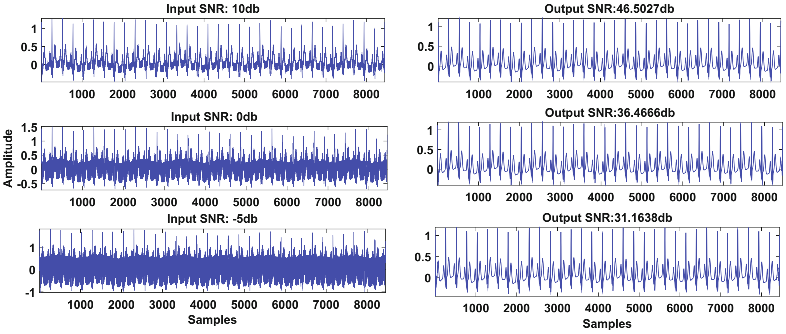


Fig. 3. Performance of the proposed approach on synthetic ECG corrupted at various noise levels.

3.2 Evaluations on Real ECG Signals

The ECG signals from MIT-BIH Arrhythmia database sampled at 360 Hz is used for experimental evaluation of the proposed approach [19]. The first 5000 samples of various ECG records namely 228, 203, 101, and 116 is considered. The experiments are carried out by introducing different noise-levels to the original signal record. From the results obtained, it is evident that the circular convolution filtering method discussed in this paper is effective for PLI removal in ECG signal. For the experimental evaluation, the noise level is varied from 28 dB to −17 dB. Table 2 shows the SNR based evaluation of the proposed method. From the experiments performed it is observed that the control parameter gives better output SNR in the range [0.001,0.04]. During the experiment it is found that the amplitude of the denoised signal reduces by a constant factor and hence the denoised signal suffers from a negative dc shift. Figure 4 shows the power spectrum density (PSD) plot of the noisy and denoised signals. In the noisy PSD plot, the presence of 60 Hz PLI is confirmed through the lobe present at 60 Hz and the denoised signal removes them efficiently.

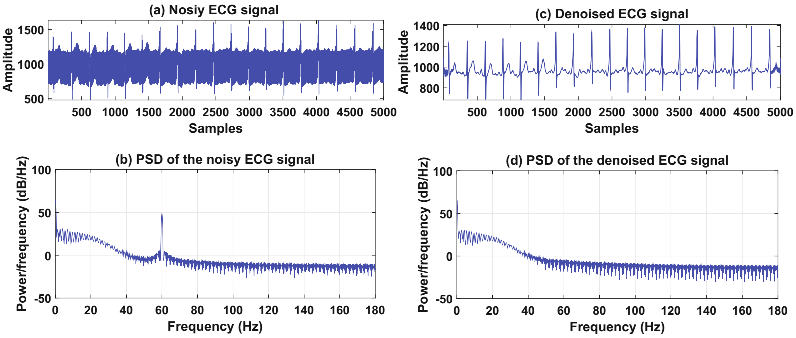


Fig. 4. (a) Noisy ECG signal (input SNR is 10 dB), (b) PSD plot of noisy ECG signal, (c) PLI removed ECG signal, (d) PSD plot of PLI removed ECG signal.

In [22], discusses a robust power-line cancellation method along with its hardware design which obtained state-of-the-art result. Since the proposed method use FFT operation, commonly included in any signal processing hardware library, the prototype of the proposed method can be fabricated easily. Table 3 shows the output SNR obtained in [22]. Comparing with results in Table 2, it can be observed that the proposed method out performs the result in [22].

Table 2. SNR based evaluation of proposed method on real ECG signals

Signal	Input SNR (dB)					
	28	10	0	−5	−10	−17
	Output SNR (dB)					
Rec- 228	55.36	46.94	41.44	39.34	36.86	33.69
Rec-203	45.21	33.71	26.70	21.80	18.17	12.67
Rec-108	47.83	34.67	25.93	21.69	17.87	12.38
Rec-116	46.6	34.33	26.12	21.90	17.90	12.61

Table 3. Output SNR from [22]

Test	Input SNR (dB)	Output SNR (dB)
Test 1	0	31.1
Test 2	−10	29.5
Test 3	0	29.2

Table 4 shows the effectiveness of the proposed method to remove the PLI even when the power-line frequency is varied as 49 Hz, 50 Hz, and 51 Hz.

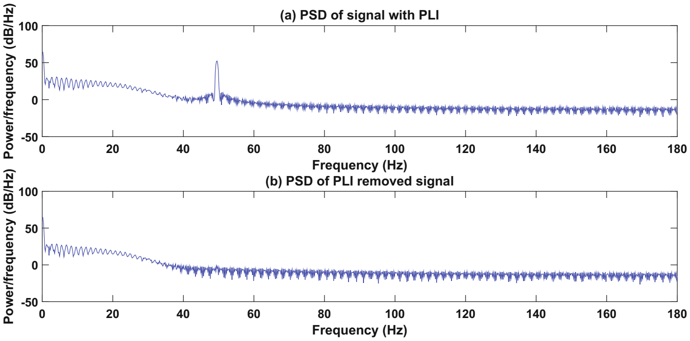


Fig. 5. (a) PSD plot of signal having power-line frequency at 49.5 Hz (input SNR is 10 dB), (b) PSD plot of PLI removed ECG signal.

The input SNR is chosen as 8 dB with $\lambda = 0.021$. This shows that the proposed method does not fail in removing PLI when the fundamental frequency deviates. Figure 5, depicts the performance on signals whose fundamental frequency deviates from 50 Hz. Figure 6 shows that this approach efficiently tracks the drifted ECG portions and removes the noise precisely.

Table 4. Output SNR for different power-line frequency having same input SNR of 8 dB

Power-line freq. (Hz)	Output SNR (dB)
49	40.40
50	41.39
51	39.36

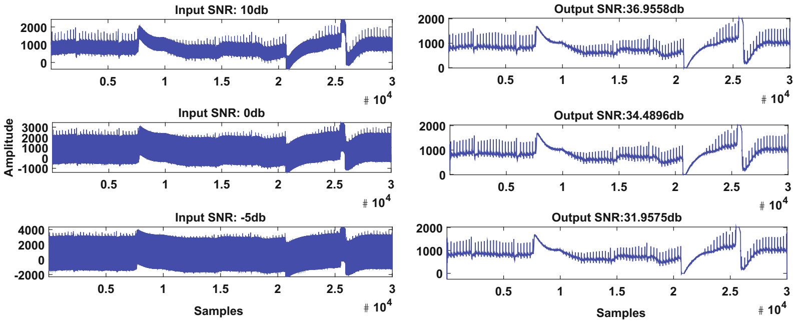


Fig. 6. Drifted ECG Signal portion from record 116 with denoising at various noise levels.

3.3 Evaluations on PCG Signals

The effectiveness of the proposed circular convolution filtering method is demonstrated in phonocardiogram (PCG) signals, which captures the heart activities [21]. PCG plays an important role to detect the presence of several cardiovascular abnormalities. The PCG signals taken from [20] is used for evaluation. Figure 7 depicts the performance on PCG signal corrupted by −1 dB noise. The high output SNR obtained suggests that the performance of the proposed method is appreciable in PCG signal denoising. The SNR value measured on various noise levels ranging from 15 dB to −5 dB is tabulated in Table 5. The parameter λ is fine tuned to 0.001 to get maximum noise reduction, indicated through high output SNR.

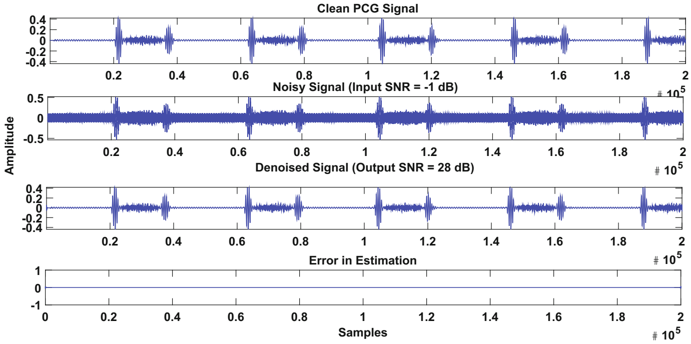


Fig. 7. Performance of the proposed approach on PCG signal corrupted by −1 dB noise (a) Clean PCG, (b) Noisy PCG, (c) Denoised PCG, (d) Error in estimation.

Table 5. SNR based evaluation of proposed method on PCG signal ($\lambda = 0.001$)

Input SNR (dB)	Output SNR (dB)
15	45.30
5	35.84
−1	28.91
−5	24.69

4 Conclusion

This paper proposes a novel circular convolution based approach for power-line interference removal from ECG signals. Since the proposed approach involves only FFT operation, it can provide a less computation intensive prototype fabrications for standalone and wearable type sensor hardware. The reason for the effectiveness of the proposed method need to be explored theoretically. The proposed method is experimentally evaluated on synthetic ECG signals and real ECG signal records from MIT-BIH Arrythmia database. The noise levels considered for the evaluation falls in the range [−17 dB, 28 dB]. It is observed from the evaluations that the proposed method obtains a state-of-the-art output SNR. Also, the performance is evaluated on PCG signals and found to be accurate.

References

1. Razzaq, N., Sheikh, S.A.A., Salman, M., Zaidi, T.: An intelligent adaptive filter for elimination of power line interference from high resolution electrocardiogram. IEEE Access **4**, 1676–1688 (2016)
2. https://en.wikipedia.org/wiki/Electrocardiography/media/File:SinusRhythmLabels.svg

3. Dobrev, D.P., Neycheva, T.D.: Automatic common mode electrode-amplifier impedance balance with SPLL synchronization. In: 2016 XXV International Scientific Conference Electronics (ET), Sozopol, pp. 1–4 (2016)
4. Sharma, T., Sharma, K.K.: Power line interference removal from ECG signals using wavelet transform based component-retrieval. In: 2016 International Conference on Advances in Computing, Communications and Informatics (ICACCI), Jaipur, pp. 95–101 (2016)
5. Meidani, M., Mashoufi, B.: Introducing new algorithms for realising an FIR filter with less hardware in order to eliminate power line interference from the ECG signal. IET Signal Process. **10**(7), 709–716 (2016)
6. Warmerdam, G.J.J., Vullings, R., Schmitt, L., van Laar, J.O.E.H., Bergmans, J.W.M.: A fixed-lag Kalman smoother to filter power line interference in electrocardiogram recordings. IEEE Trans. Biomed. Eng. **PP**(99), 1 (2016)
7. Hu, X., Xiao, Z., Liu, C.: Reduction arithmetic for power line interference from ECG based on estimating sinusoidal parameters. In: 2010 3rd International Conference on Biomedical Engineering and Informatics, Yantai, pp. 2089–2092 (2010)
8. Vijendra, V., Kulkarni, M.: ECG signal filtering using DWT haar wavelets coefficient techniques. In: 2016 International Conference on Emerging Trends in Engineering, Technology and Science (ICETETS), Pudukkottai, pp. 1–6 (2016)
9. Tomasini, M., Benatti, S., Milosevic, B., Farella, E., Benini, L.: Power line interference removal for high-quality continuous biosignal monitoring with low-power wearable devices. IEEE Sens. J. **16**(10), 3887–3895 (2016)
10. Satija, U., Ramkumar, B., Manikandan, M.S.: Low-complexity detection and classification of ECG noises for automated ECG analysis system. In: 2016 International Conference on Signal Processing and Communications (SPCOM), Bangalore, pp. 1–5 (2016)
11. Ahlstrom, M.L., Tompkins, W.J.: Digital filters for real-time ECG signal processing using microprocessors. IEEE Trans. Biomed. Eng. **9**, 708–713 (1985)
12. Hamilton, P.S.: A comparison of adaptive and nonadaptive filters for reduction of power line interference in the ECG. IEEE Trans. Biomed. Eng. **43**, 105–109 (1996)
13. Mishra, S., Das, D., Kumar, R., Sumathi, P.: A Power-Line Interference Canceler Based on Sliding DFT Phase Locking Scheme for ECG Signals. IEEE Trans. Instrum. Meas. **64**, 132–142 (2015)
14. Avendano-Valencia, L.D., Avendano, L.E., Ferrero, J.M., Castellanos-Dominguez, G: Improvement of an extended Kalman lter power line interference suppressor for ECG signals. IEEE Comput. Cardiol., pp. 553–556 (2007)
15. Bharath, H.N., Prabhu, K.M.M.: A new LMS based adaptive interference canceller for ECG power line removal. In: 2012 International Conference on Biomedical Engineering (ICoBE), pp. 68–73 (2012)
16. Mohan, N., Kumar, S., Poornachandran, P., Soman, K.P.: Modied variational mode decomposition for power line interference removal in ECG signals. Int. J. Electric. Comput. Eng. (IJECE) **6** (2016)
17. Soman, K.P., Ramanathan, R.: Digital signal and image processing-the sparse way. Isa Publication (2012)
18. McSharry, P.E., Clifford, G.D., Tarassenko, L., Smith, L.A.: A dynamical model for generating synthetic electrocardiogram signals. IEEE Trans. Biomed. Eng. **50**(3), 289–294 (2003)
19. https://physionet.org/physiobank/database/mitdb/. Cited 22 March

20. http://www.med.umich.edu/lrc/psbopen/html/repo/primerheartsound.html. Cited 22 March
21. Manikandan, M.S., Soman, K.P.: Robust heart sound activity detection in noisy environments. Electron. Lett. **46**(16), 1100–1102 (2010)
22. Keshtkaran, M.R., Yang, Z.: A fast, robust algorithm for power line interference cancellation in neural recording. J. Neural Eng. **11**, 026017 (2014)

Object Detection and Localization Using Compressed Sensing

Poonam Ashok Deotale(✉) and Preetida Vinayakray-Jani

Department of Electronics and Telecommunication Sardar Patel
Institute of Technology, Mumbai University, Mumbai, India
poonam.deotale@gmail.com, preeti.vinayakray@gmail.com

Abstract. Localization is very significant in Underwater Sensor Network (UWSN) applications. The functionality of the network can face challenges by the force of water current and hostile environmental conditions. This work presents a method for localization using Compressed Sensing (CS). It is implemented without using GPS technology which makes the method reliable. CS is employed in the data acquisition module for transmission and reconstruction of audio signal. It is a dictionary based execution exploiting l_1 minimization using Gabor transform. Here, localization using audio is performed using Time Difference of Arrival (TDOA). Moore-Penrose pseudo-inverse is used for matrix operations. An array of audio sensors or hydrophones is assumed while performing this work. The results of simulation indicate that this is an efficient technique for object detection and localization.

Keywords: Compressed sensing · Localization · TDOA · Underwater sensor network

1 Introduction

Locating the objects that lie on the ocean floor or detecting the presence of underwater animals is significant for many applications in UWSN. Amongst the many technologies that have been discussed in research studies, most of them rely on GPS function. However, one cannot always depend upon GPS technology [1]. GPS works very efficiently in the terrestrial environment but there are some characteristics of underwater nature that does not allow it to function so well there [30]. For any electromagnetic wave, travelling from one medium to another medium possessing different conductivity, permittivity or permeability triggers some amount of reflection. The skin depth of any wave depends on the frequency and conductivity. This makes it difficult for waves with high frequencies to survive underwater. GPS signals are of high frequencies around 1.3 GHz. That is why GPS cannot penetrate water medium successfully.

Data acquisition and transmission is a challenging task in UWSN due to its limited bandwidth. As explained above, signals with high frequencies cannot function well here. Besides; various technical issues like multipath propagation, absorption and scattering gets heightened due to the conductivity of water.

S.M. Thampi et al. (eds.), *Advances in Signal Processing and Intelligent Recognition Systems*, Advances in Intelligent Systems and Computing 678,
DOI 10.1007/978-3-319-67934-1_12

The hardware that is setup underwater tends to shift with the water currents [29]. Hence, it is essential to devise a localization technique that can function faster within the limited bandwidth that is available. Compressed sensing is used to enhance the data acquisition system in UWSN because it utilizes minimum number of transmission and hence less time. It can produce output with less investment of time and energy. CS has been famously used to tackle a huge exemplar of data in the shortest time interval. However, the traditional methods of CS make computations very complex if we use fixed data representations for reconstructing data. This work uses dictionary learning for data de-noising and reconstruction.

The rest of the paper is structured as follows: Sect. 2 throws light on the literature survey done in the process of this work. The sub-section 2.1 describes the salient features of SONAR by providing its high level classification while in Sect. 2.2, the theory of compressed sensing is explained by giving one example. Section 3 gives the complete walk through on the experimental implementation. Section 4 discusses the results of the simulation. Section 5 concludes this paper giving a note on the future work. Section 6 acknowledges the resources provided for this research work.

2 Related Work

In the past, a lot of research work has been published with regards to localization in terrestrial as well as underwater sensor networks. In [1], the localization methodologies have been categorized into target based techniques and self-localization techniques (like range-based and range-free methods). It says that the target or source based methods can be applied in underwater networks for the purpose of locating aquatic animals and sunken ships. However, it does not give the exact approach which is to be followed to survive the underwater constraints. The work demonstrated in [2] is a practical implementation of UWSN by a mobile anchor node and 4 non-coplanar nodes. Here, localization is realized by position information of the 4 non-coplanar nodes transmitted to the anchor. It relies on the information transmitted by the peer nodes to the anchor. In [6], an acoustic monitoring system is presented which detects the source by comparing with a known audio clip. The detection of the target is clearly based on an audio signal known a priori. However, the signal that one may come across cannot be predicted and hence we cannot depend upon pre-defined values The work done in [3–5] uses SONAR images for detection and localization of underwater objects. [5] uses the blob detection technique to identify the peak response location where the certainty of object being located is predicted. This process may prove inefficient if multiple instances of peak are found. Simultaneous use of sound and image has been described in [7] and [8]. However, it is identical to smart-home applications and also the sound and image data are processed separately with no common linkage [10,11]. Gives a detailed study on the application of compressed sensing to speech and audio signals. The work in [13] visits myriad realms of applications of sparse signals and gives a deep breadth for compressed sensing techniques. In [14], a set of 3 hydrophones forms a virtual system where the localization function is based on GPS which cannot be trusted as explained in Sect. 1 above. This paper executes localization independent of any electromagnetic signals like GPS. The material in [17] is a detailed walk through various mechanisms of positioning in Wireless Sensor Networks (WSNs). The work in [12] has used TDOA method but without CS. It compares the parameterization method for TDOA with cross-correlation. The main inspiration for this work was derived from [9] where the localization of the target is achieved by TDOA using compressed sensing. Though it reduces number of transmissions through CS, [9] uses fixed data representations which may give erroneous results for indeterminate signals like speech signals. To avoid this, we have implemented Short Time Fourier Transform (STFT) in a Gabor Dictionary.

This work assumes a clustered network of audio sensors placed underwater. Localization of mobile target is the objective here which brings us to clustering techniques. The work in [22] shows how "parallel genetic algorithms" can be used to optimize the open-shop scheduling algorithm by implementing Message Passing Interface (MPI) on a Beowulf cluster. The technique proposed here can be applied in processes involving numerous tasks/jobs adhering to a pre-defined schedule. However, the present problem relies on the target detection which is of unpredictable type. The research in [24] proposes a novel multi-stage "Adaptive Charged System Search (ACSS) algorithm" for optimal tuning of fuzzy controllers. It is an optimization of search algorithms like [22] and can be applicable in embedded systems. In [25], a hybrid approach is put forth by making use of both local and global search methods for parameter tuning. In [23], Inner Product Induced Norm based Consistent Dissimilarity (IPINCD) measure is defined which throws optimal solution in case of clustering techniques where convex functions are employed. The clustering methodology can be optimized using this measure. The application of clustering in TDOA based localization methods is explained in [26,27]. The estimation of the position of a node in a WSN whether distributed or clustered is demonstrated as a convex problem in [28]. The study of [28] helped us realize the possibility of application of IPIN in this work.

2.1 SONAR

Sonar (Sound Navigation And Ranging) represents the technological paradigm that serves applications requiring location estimation in underwater environment. From [19–21], we have derived findings on the phenomena of SONAR sensing. There are two types of sonar systems - active and passive.

Active Sonar. Inside water, the distance of the object can be calculated by tracking the time elapsed between the sound signal thrown and the signal reflected back from the target object. This logic is applied in SONAR systems. SONARs can be side-scan sonar, multi-beam sonars, etc. These devices are capable of emitting highly directional pings of acoustic signal. Active sonars are capable of not only receiving sound signals but also generating deliberate acoustic signals [19].

Passive Sonar Systems. Passive SONAR systems are nothing but listeners. They are capable of detecting sound signals emitted by underwater animals or ships or sound generated by any underwater activity. For example, an enemy submarine that is passing by the vicinity can be detected by detecting the sound emitted by its engines, etc. Unlike active sonars, passive sonars do not generate any signals which is expected to be echoed. Passive SONARs can be an array of hydrophones or acoustic sensors. We have used these characteristics in our algorithm for object detection and subsequent localization [31].

2.2 Compressed Sensing

Nyquist-Shannon principle states that a signal can be sampled and restored accurately by sampling it with at least twice its original frequency. But what if the source signal is distorted due to white noise and we miss a few bytes of the original signal as explained in [16]. It should also be taken into account that many interfering signals may not be white noise. These signal can be sparse, for instance. Let us consider a time-domain signal 'x' of length N. Following conventional methods, we would take N measurements for sampling purpose. However, if we are to use CS we need to seek only K number of measurements $(K < N)$. Let us denote these measurements by the vector y. In real life, almost all the signals are non-linear. These signals cannot be realized by picking one measurement for one input value. As a result, to acquire y, we make use of y = Ax where A is the sensing matrix. Here, A is nothing but the basis of the vectors which help us to get to y from x. It can be a simple identity matrix, a transform or any non-linear operator [16].

We need to keep in mind that the basis functions need to be chosen such that it falls suitable to the desired application. Consider ψ as the celebrated Discrete Fourier Transform (DFT). We take cognizance of the fact that the sparsity is the pre-requisite for compressed sampling or compressed sensing for that matter. Hence, suppose the input is a sparse signal 'c'. So, ψ becomes the sensing basis. Its function is to act as the representation domain through which we extract the signal values. $b = \phi$f is a linear operator that samples our signal. Thus, we have a time domain input signal with DFT sensing basis which converts it to frequency domain as described in [16]. We must solve Ax = b, where A = $\varPhi\varPsi$ in order to recover the signal coefficients. Thereafter, we need to compute f $\approx \varPsi x$ to recover the signal. A has more columns than rows because the procedure is that of compression. Computation of x leads us to the fact that there are more number of unknowns than equations. This is when the role of the l_1 norm regularization comes into picture. Figure 1 gives the above discussed example.

2.3 Why l_1 Norm for Sparsity

l_1 norm and l_2 norm are both used for bringing about regularization [16]. We have chosen l_1 norm because of its certain properties. Suppose there is a vector $\vec{x} = (1, \epsilon) \in \mathbb{R}^2$ where $\epsilon > 0$ is very small. We can produce l_1 and l_2 norms of $\vec{x}$ by following:

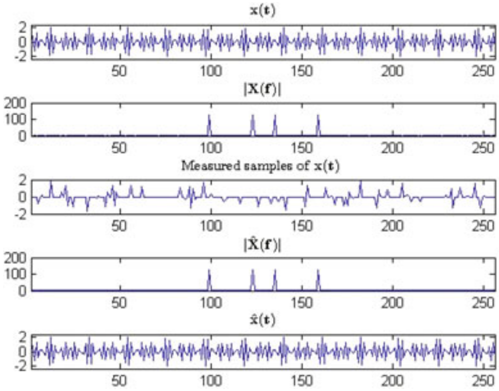


Fig. 1. A sparse frequency-domain signal

$$||\vec{x}||_1 = 1 + \epsilon - l_1 Norm \tag{1}$$

$$||\vec{x}||_2^2 = 1 + \epsilon^2 - l_2 Norm \tag{2}$$

For regularization to materialize, the magnitude of one of the elements of **x** should be reduced by $\delta \leq \epsilon$. So, we change x_1 to 1 - δ where delta is a value by which we are reducing the magnitude of one of the elements (x_1). Thus, the resulting norms are:

$$||\mathbf{x} - (\delta, 0)||_1 = 1 - \delta + \epsilon - l_1 Norm \tag{3}$$

$$||\mathbf{x} - (\delta, 0)||_2^2 = 1 - 2\delta + \delta^2 + \epsilon^2 - l_2 Norm \tag{4}$$

If we decrease x_2 by δ , it gives us:

$$||\vec{x} - (0, \delta)||_1 = 1 - \delta + \epsilon - l_1 Norm \tag{5}$$

$$||\vec{x} - (0, \delta)||_2^2 = 1 - 2\epsilon\delta + \delta^2 + \epsilon^2 - l_2 Norm \tag{6}$$

For l_2 penalty, if we regularize the larger term x_1, it produces more reduction in norm than implementing the same in the smaller term $x_2 \approx 0$. On the contrary, for l_1 penalty; the reduction amounts to the same result. We can thus deduce that penalizing an exemplar with l_2 norm does not necessarily provide solutions of zero value. This is due to the fact that reduction in l_2 norm from ϵ to 0 is almost nonexistent when ϵ is so small. Whereas, penalizing by l_1 norm almost every time equates towards δ irrespective of what model is being penalized. This paper employs l_1 penalization is for compressing the audio signal [15].

3 Methodology

The method described in this paper is a continuation of the work explained in [18]. In order to implement CS for audio signals, the work involved computing a set of Short Time Fourier transforms (STFTs) by sliding windows and provided to the Gabor Transform. Using Basis pursuit, which optimizes l_1 minimized co-efficients; this paper implements dictionary to de-noise an audio signal.

3.1 Virtual System

The work efficiently applies CS for detection and localization of objects using audio. The CS technique was upgraded to be compatible with any audio signal. To implement this method, we have used the dolphins whistle as the data set. The data was obtained from 'soundbible.com' website; with the knowledge that the output of a passive SONAR can be obtained in WAV format [32]. The architecture of the proposed system is illustrated in Fig. 2. We have assumed an array of seven hydrophones. A single cluster head is dedicated to a fixed no. of sensor nodes whose location is known a priori. This work presents the involvement of CS in enhancing data acquisition system by reducing its computational complexity and offering reliability. This paper puts forth an optimizing algorithm based on "Clustering using Inner Product Induced Norm(IPIN) based dissimilarity measures" [23]. The entire system is summarized by the flowchart diagram in Fig. 3.

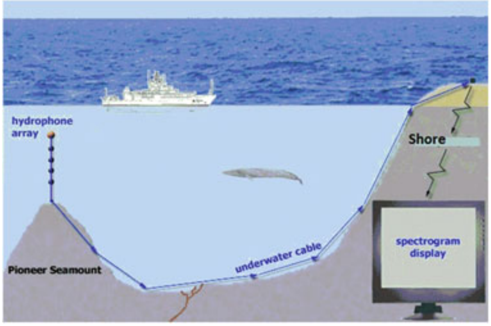


Fig. 2. Underwater sensor architecture

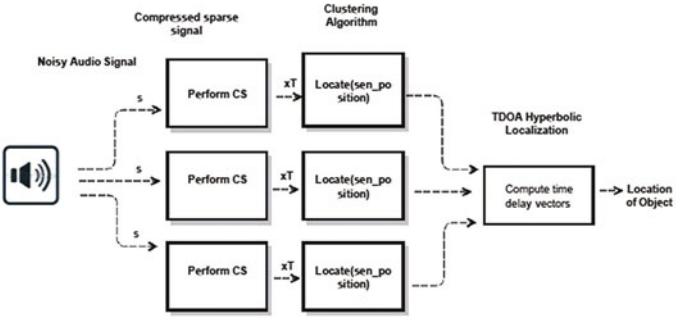


Fig. 3. Localization using clustering

3.2 Compressed Sensing of Audio

To solve the problem of localization by audio signals, method is developed so as to serve unpredictable signals like speech signals. As the audio is expected to come from aquatic animals, the signal parameters cannot be easily predicted. The traditional methods like FFT, DFT, etc. relying on static waveform variations are not sufficient in providing the required information. The nature of speech signals varies with time as well as frequency. STFT performs well in cases where there is variable frequency surpassing different time locations. However, it provides solely fixed resolution. Hence the work proposed here implements window functions over the derived STFTs. This is exploited in the Gabor Transform. The flow for localization using audio signal is explained in detail in Fig. 4.

Algorithm 1. Localization Using Audio Signal

```
Reception of the signals by M passive SONARs or hydrophones
Detect the noisy input audio signal ‘s’ and load the sound
Initialize the number and size of windows ‘L”
Establish dictionary redundancy
Compute STFT with tight frame xT
Execute Gabor transform
De-noise Audio
Establish convex constraints and formulate h(x)
Clustering Algorithm
TDOA localization
Audio reconstruction
```

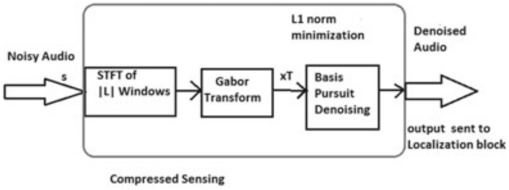


Fig. 4. Compressed sensing for audio data

3.3 Cluster Localization by Convex Optimization Algorithm

The audio sensors or hydrophones or passive SONARs are assumed to be deployed in the UWSN in the form of a cluster or a group of clusters. Cluster localization is implemented using convex optimization algorithm. Using Time Distance of Arrival (TDOA), the distance of the object from the sensors is calculated and the object is located. This work proposes a localization method using convex optimization clustering algorithm by applying Clustering with

Inner Product Induced Norm (IPIN) based Dissimilarity Measures as suggested by Arkajyoti Saha and Swagatam Das in [23]. It helps us in attaining formidable clustering because it has been proved to provide natural resistance against noise and errors caused due to mobility of target [28]. The fundamental class of dissimilarity measures on which this algorithm is realized is defined below:

"A function $dist : \mathcal{M}^d X \mathbb{R}^d X \mathbb{R}^d \rightarrow \mathbb{R}_+$ is called an Inner Product Induced Norm based Consistent Dissimilarity (IPINCD) measure with respect to some convex set

$C_1 \subseteq \mathbb{R}_d$ and $C_2 \subseteq \mathcal{M}^d$ if for some function h: $\mathbb{R} \rightarrow \mathbb{R}_+$, $dist(\mathbf{M},\mathbf{y},\mathbf{x}) = d_M(\mathbf{y},\mathbf{x}) = h((x-y)^T M(x-y))$ where the following conditions hold:

1. h is differentiable on $\mathbb{R}_+$.
2. $\mathbf{M} \rightarrow dist(\mathbf{M},\mathbf{y},\mathbf{x})$ is a convex function on $C_2, \forall \mathbf{y} \in \mathbf{C}_1$ and $\mathbf{y} \rightarrow dist(\mathbf{M},\mathbf{y},\mathbf{x})$ is strongly convex function on $\mathbf{C}_1, \forall \mathbf{M} \in \mathbf{C}_2$.
3. $\mathbf{M}\frac{\partial dist(\mathbf{M},\mathbf{y},\mathbf{x})}{\partial \mathbf{M}}$ and $\mathbf{y} \rightarrow \frac{\partial dist(\mathbf{M},\mathbf{y},\mathbf{x})}{\partial \mathbf{y}}$ are Lipschitz continuous functions (on $\mathbf{C}_1$ and $\mathbf{C}_2$ respectively) $\forall \mathbf{y} \in \mathbf{C}_1$ and $\forall \mathbf{M} \in \mathbf{C}_2$ respectively."

In the above definition, *dist* is a symmetric "family of functions". The first 2 points are meant for the convergence analysis of this algorithm while the last one leads to convergence analysis related to optimization of cluster representatives. One important observation is that for condition 2, convex property of the function 'h' is sufficient.

When the above algorithm is applied as exponential IPIN based dissimilarity measure, then it is useful in developing a robust clustering technique according to [23]. We propose implementing this formulation in localization of the cluster in our sensor network. The measure can be realized as below:

$$dist(\mathbf{M}, \mathbf{y}, \mathbf{x}) = exp((x-y)^T \mathbf{M}(x-y)), \forall \mathbf{M} \in \mathcal{M}^d, y, x \in \mathbb{R}^d. \tag{7}$$

Thus, 'h' here which is presumed to be convex function, is also exponential now. What remains is to establish that h(x) where $x \in \mathbb{R}$ is the distance estimation function for our cluster network and is convex at the same time in order for above principle to be applicable to it. Estimating the position of a node involves establishing several convex models as well explained by Doherty et al. in [28]. Thus, the sufficient conditions are met in order to apply the clustering optimization algorithm for localization. At present we are working on optimizing our simulation program using this algorithm.

3.4 TDOA Implementation

The conventional method of Time of Arrival (TOA) can give a direct path to locate elements from distance measurements but it also has a few disadvantages. For this method to function, all the participating nodes must have their nodal clocks accurately synchronized. When the data is being transmitted from the transmitter to receiver, the time when the transmission was initiated should be appropriately communicated. Similar to TOA, Time Distance of Arrival (TDOA) is also a geometric method [12]. However, it does not face the above mentioned short comings. It uses the difference in the time at which the desired object's sound is received at the different audio sensors. The time at which it actually reaches the sensing device is ignored here. But simply a single value of the time difference is not sufficient to estimate the co-ordinates of the object. Hence, it requires a minimum of three sensors to triangulate the geometric position. From our experiments we realize that the location estimation accuracy is sufficiently achieved with a minimum of 5 audio sensors. In our simulation, we have used 7 hydrophones. Figure 5 here portrays the basics of TDOA. In our simulation, we have used cross-correlation method. F1 and F2 are the reference positions of the 2 hydrophones while the mobile target is T. Let t_0 be the time instance at which the sound signal is transmitted from T. Figure 6 shows the hyperbolic trajectory with F1 and F2 at focal positions. At F1, it is received at t_1 and at F2, it is received at t_2. Figure 7 depicts the timing signals of a pulse transmitted from T at time t_0, received at F1 at time t_1 and at F2 at time t_2. Clocks of F1 and F2 are synchronized but not of T; so t_0 is unknown. However, the time difference of arrival at F1 and F2 can be calculated as originally written in [12]:

$$t_1 - t_2 = (t_1 - t_0) - (t_2 - t_0) = \tau_m \tag{8}$$

Thus, if we take 'v' to be the velocity of the medium, the total distance between the reference stations and the target is found to be:

$$\Delta d = d_1 - d_2 = v * (t_1 - t_2) \tag{9}$$

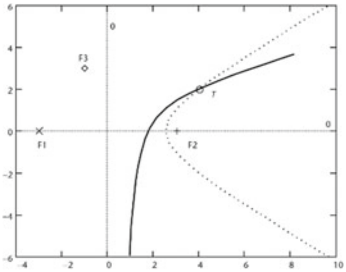


Fig. 5. TDOA localization using 3 fixed signals

The locus of the points (reference stations) from which Δd should be constant forms a hyperbola. The desired mobile target is estimated to be somewhere on this hyperbola. As the measurements from 2 stations amounts to a proximity anywhere on an hyperbola, a third station is required to denote the desired target as shown in Fig. 5.

For the purpose of localization, we have drawn a virtual system where there are M number of hydrophones in a distributed network. Here, $f(xm; ym; zm)_m^M$ are the co-ordinates of these hydrophones which are nothing but audio sensors

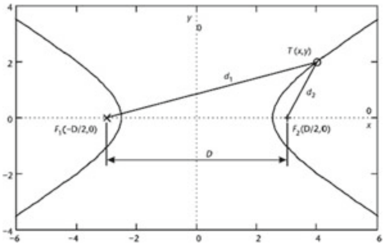


Fig. 6. Geometry of TDOA localization

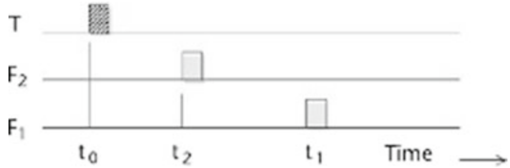


Fig. 7. Timing signals

underwater. The co-ordinates of the unknown target source are (x; y; z). Let t_{F1} be the time of transit from the source i.e. T to hydrophone F1. So, $t_1 - t_0 = t_{F1}$. The value of 'v' is 340.29 m/s in air and 1484 m/s in water. Let $R_{F1} = v * T_{F1}$ be the distance between the source and the hydrophone F1. Following equation needs to be solved to estimate the distance R_{F1} between the source and sensor:

$$R_{F1}^2 = (v * \tau_m + R_{F2})^2 = v^2 * \tau_m + 2 * v * \tau_m * R_{F2} + R_{F2}^2 \tag{10}$$

where R_{F2} is the distance between the source T and the hydrophone F2.

The equation is solved using matrix operations and executed in MATLAB by applying Moore-Penrose pseudo-inverse. It is implemented for the underwater as well as terrestrial environments.

4 Results and Analysis

As we are considering the underwater environment, the incoming audio should be noisy. The sound is initially aggregated with Gaussian white noise. The reconstruction of the signal was executed by soft thresholding and by dictionary learning. The problem with soft thresholding technique is that the threshold should be set correctly which is quite difficult because of the redundancy of the representation. Hence it is concluded that dictionary learning is the right approach. The process executed employs l_1 norm minimization which can be formulated as: $min_{S_1} 1/2 * norm(x - x_1)^2 + \lambda * norm(S_1, 1)(*)$.

The above equation gives the output $|S1|$ which is a set of Gabor coefficients. The signal reconstructed from these coefficients is $|x1|$. The quality of audio reconstruction by using Gabor Transform is evaluated by the computing SNR.

We took three instances of reconstruction and the SNR values are equal in both the methods which we have placed here in Figs. 8 and 9 respectively. The average SNR comes out to be 25.2. Approximate redundancy of the dictionary = 8. True redundancy of the dictionary = 8.0586. Reconstruction error (should be 0) = 1.58e-16. Elapsed time is 6.120767 s. We can thus infer that our theory is feasible. In order to evaluate the Localization technique; we implemented the localization with (Fig. 10) and without (Fig. 11) optimization using CS. The comparison of true position and estimated position in Figs. 10 and 11 say that by using CS more accurate results are obtained than by the traditional sampling method. Also, we compared our results in terms of computational speed. Our results says that the elapsed time by applying CS is 6.120767 s. The same without optimization takes 22.355459 s.

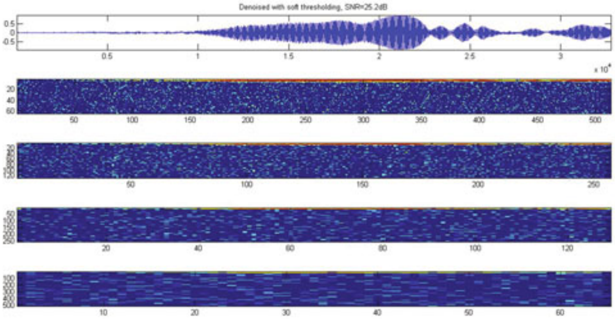


Fig. 8. SNR obtained by dictionary learning using soft thresholding

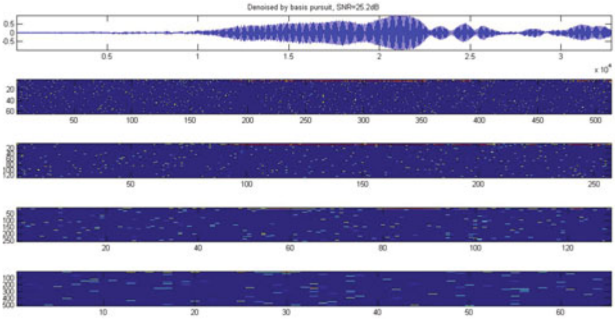


Fig. 9. SNR obtained by dictionary learning using basis pursuit

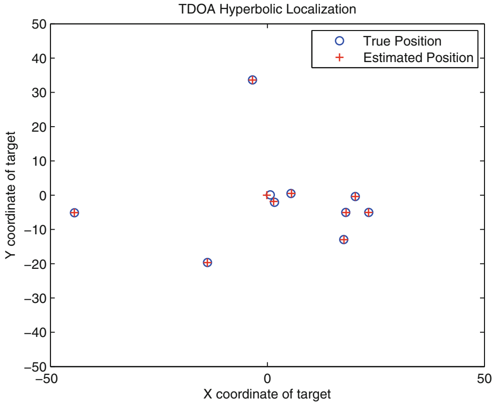


Fig. 10. Localization without CS

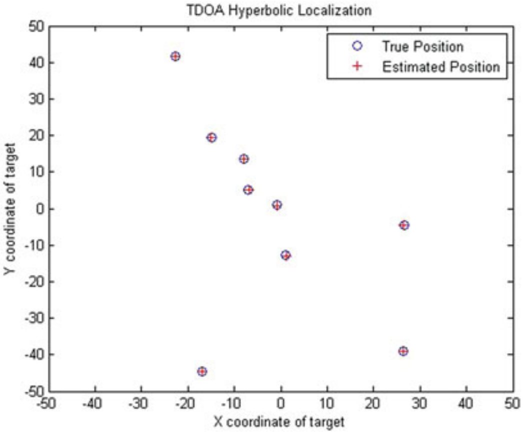


Fig. 11. Localization with CS

5 Conclusions and Future Scope

The work shown in this paper gives a brief description of a novel methodology to detect and localize objects using CS with the target network being Underwater Sensor networks. The system performs sparse reconstruction of data. The sparsity function relies on l_1 minimization. CS of audio was executed by using a set of STFTs computed with multiple windows as an input to the Gabor Transform. Using Basis pursuit, this paper has implemented dictionary for the audio signal. The methods used for acquiring dictionary reduces the computational complexity which means that our method not only guarantees reliability but also energy efficiency. The concept of sparsity is applied here for denoising data. Detection of objects follows after the data is reconstructed. Since the data acquisition system exhibits energy efficiency, it can be applied to numerous scenarios like rescue expeditions in inaccessible areas. The experimental results show that the reconstruction of data at receiver's end is obtained with considerable reduction in complexity with this approach.

The localization using audio signal is executed by following the TDOA principle. Though the implementation is based on simulation, it can be realized in practical environment. Future work will involve trying this method in a real-time acoustic sensor network. This paper proposes an optimization algorithm (IPINCD) for localization based on clustering to be implemented in future.

Acknowledgments. The authors would like to thank Sardar Patel Institute of Technology, India for providing the necessary facilities for carrying out this work.

References

1. Cheng, L., Wu, C., Zhang, Y., Wu, H., Li, M., Maple, C.: A survey of localization in wireless sensor network. Int. J. Distrib. Sens. Netw. 1–5 (2012). Hindawi Publishing Corporation
2. Nuo, W., Ming-Lei, S., Ming, Y., Yuanyuan, Y., Jiyong, X.: A localization algorithm for underwater wireless sensor networks with the surface deployed mobile anchor node. In: Sixth International Conference on Intelligent Systems Design and Engineering Applications, pp. 30–32 (2015)
3. Weng, L., Li, M., Gong, Z.: On sonar image processing techniques for detection and localization of underwater objects. Appl. Mech. Mater. **236–237**, 509–514 (2012)
4. Zao, S.: Automatic underwater multiple objects detection and tracking using sonar imaging. ACM Digital Library, pp. 14–35 (2010)
5. Akshaya, B., Narmadha, V., Sree Sharmila, T., Rajendran, V.: Sparse representation to localize objects in underwater acoustic images. In: IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT), pp. 1–5. Coimbatore, India, August 2015
6. Ismail, M.F.F.B., Yie, L.W.: Acoustic monitoring system using wireless sensor networks. Procedia Eng. **41**, 68–74 (2012). In: International Symposium on Robotics and Intelligent Sensors 2012 (IRIS 2012), Science Direct
7. Suzuki, T., Kato, K., Makihara, E., Kobayashi, T., Kono, H., Sawai, K., Kawabata, K., Takemura, F., Isomura, N., Yamashiro, H.: Development of underwater monitoring wireless sensor network to support coral reef observation. Int. J. Distrib. Sens. Netw. **2014**, 1–10 (2014). Hindawi Publishing Corporation, Article ID: 189643
8. Tavakoli, A., Pourmohammad, A.: Image denoising based on compressed sensing. Int. J. Comput. Theor. Eng. **4**(2), 266–269 (2012)
9. Jiang, H., Mathews, B., Wilford, P.: Sound localization using compressed sensing. In: Proceedings of SENSORNETS, pp. 159–166 (2012)
10. Christensen, M.G., Ostergaard, J., Jensen, S.H.: On compressed sending and its applications to speech and audio signals. In: Conference Record of the Forty-Third Asilomar Conference on Signals, Systems and Computers (2009)
11. Griffin, A., Tsakalides, P.: Compressed sensing of audio signals using multiple sensors. In: 16th European Signal Processing Conference (2008)
12. Dong, Z., Yu, M.: Research on TDOA based microphone array acoustic localization. In: IEEE 12th International Conference on Electronic Measurement and Instruments (2015)

13. Peyre, G., Fadili, J.M.: Learning analysis sparsity priors. In: Sampta11, Singapore, 4 pp. (2011)
14. Postolache, O., Pereira, M.D., Girao, P.: Intelligent distributed virtual system for underwater acoustic source localization and sounds classification. In: IEEE International Workshop on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications, 6–8 September 2007 (2007)
15. Bektas, S., Sisman, Y.: The comparison of L1 and L2-norm minimization methods. Int. J. Phys. Sci. **5**, 1721–1727 (2010)
16. Cands, E.J., Wakin, M.B.: An introduction to compressive sampling. IEEE Sig. Process. Mag. **25**, 21–30 (2008)
17. Bensky, A.: Wireless Positioning Technologies and Applications, 2nd edn. Artech House book (2016)
18. Deotale, P.A., Jani, P.V.: Compressed sensing for object detection using dictionary learning. In: International Conference on Computational Intelligence in Data Science, June 2017
19. Baldacci, A., Haralabus, G.: Signal processing for an active sonar system suitable for advanced sensor technology applications and environmental adaptation schemes. In: EUSIPCO 2006, European Signal Processing Conference, Florence, Italy, 4–8 September 2006 (2006)
20. Mandic, F., Rendulic, I., Miskovic, N., NaZ, Y.: Underwater object tracking using sonar and USBL measurements. J. Sens. **2016**, 10 pages (2016). Hindawi Publishing Corporation
21. Weng, L., Li, M., Gong, Z.: On sonar image processing techniques for detection and localization of underwater objects. Appl. Mech. Mater. **236–237**(2012), 509–514 (2012)
22. Ghosn, S.B., Drouby, F., Harmanani, H.M.: A parallel genetic algorithm for the open-shop scheduling problem using deterministic and random moves. Int. J. Artif. Intell. **14**(1), 130–144 (2016)
23. Saha, A., Das, S.: Optimizing cluster structures with inner product induced norm based dissimilarity measures theoretical development and convergence analysis. Inf. Sci. **372**, 796–814 (2016)
24. Precup, R.-E., David, R.-C., Petriu, E.M., Preitl, S., Radac, M.-B.: Novel adaptive charged system search algorithm for optimal tuning of fuzzy controllers. Expert Syst. Appl. **41**, 1168–1175 (2014). Elsevier
25. Johanyak, Z.C., Papp, O.: A hybrid algorithm for parameter tuning in fuzzy model identification. Acta Polytech. Hung. **9**(6), 153–165 (2012)
26. Mesmoudi, A., Feham, M., Labraoui, N.: Wireless sensor networks localization algorithms: a comprehensive survey. Int. J. Comput. Netw. Commun. (IJCNC) **5**(6), 45–64 (2013)
27. Moore, D., Leonard, J., Rus, D., Teller, S.: Robust distributed network localization with noisy range measurements. In: SenSys 2004 Proceedings of the 2nd International Conference on Embedded Networked Sensor Systems, 3–5 November 2004 (2004)
28. Doherty, L., Pister, K.S.J., El Ghaoui, L.: Convex position estimation inWireless sensor networks. In: INFOCOM 2001, Twentieth Annual Joint Conference of the IEEE Computer and Communications Societies, Proceedings. IEEE, 07 August 2002
29. Han, G., Jiang, J., Shu, L., Xu, Y., Wang, F.: Localization algorithms of underwater wireless sensor networks: a survey. Sensors (Basel) **12**, 2026–2061 (2012)

30. Dong, B., Mahdy, A.M.: GPS-free localization schemes for underwater wireless sensor networks. In: Third International Conference on Mobile Ubiquitous Computing, Systems, Services and Technologies (2009)
31. Maranda, B.H.: Passive sonar. In: Handbook of Signal Processing in Acoustics, pp. 1757–1781. Springer (2008)
32. Zora, M., Buscaino, G., Buscaino, C., D'Anca, F., Mazzola, S.: Acoustic signals monitoring in shallow marine waters: technological progress for scientific data acquisition. Proc. Earth Planet. Sci. **4**, 80–92 (2011). SciVerse ScienceDirect

Vehicle License Plate Detection Using Image Segmentation and Morphological Image Processing

Wasif Shafaet Chowdhury(✉), Ashikur Rashid Khan, and Jia Uddin

BRAC University, 66 Bir Uttam AK Khandakar Road, Dhaka 1212, Bangladesh
wasifshafaet599@gmail.com, ashikkhan.bu@gmail.com, engrjiauddin@gmail.com

Abstract. This paper presents an image segmentation technique to segment out the Region of Interest (ROI) from an image, in this study, the ROI is the vehicle license plate. In order to successfully detect the license plate an improvised Sliding Concentric Window (SCW) algorithm has been developed to perform the segmentation process. In this proposed model, vehicle images were obtained and the SCW algorithm has been performed to segment out the ROI and then Morphological Image Processing techniques named erosion and dilation have been used to locate the license plate. In order to validate our proposed model, we have used a dataset where the images of the vehicles have been taken from a different angle that contains natural background and different lighting conditions. It has been observed that the proposed model exhibits 86.5% accuracy rate for our tested dataset. In addition to that, a comparative study has been carried out between two different techniques (Improved SCW and Modified Bernsen Algorithm) of ROI detection to illustrate their accuracy rate. It has been found that the accuracy rate of the proposed model of VLP detection is higher than some other traditional algorithms.

Keywords: Image segmentation · Vehicle license plate detection · Morphological image processing

1 Introduction

Nowadays the density of traffic is increasing rapidly. In order to monitor the traffic, the traditional methods sometimes fail to provide an optimal solution [1]. In order to overcome this problem, an automatic vehicle number plate detection technique should be developed. Automatic VLP detection systems provide more effective and technical advantages than the traditional traffic monitoring systems [1]. Moreover, additional traffic information can be obtained from images including vehicle classification, line changes and a single camera can monitor multiple lines and can simultaneously read information about the vehicle [2, 3]. A number of methods regarding VLP detection have been introduced so far, but optimization and increasing their accuracy rate is a must. The reason behind, the accuracy rate of some license plate detection algorithms

S.M. Thampi et al. (eds.), *Advances in Signal Processing and Intelligent Recognition Systems*, Advances in Intelligent Systems and Computing 678,
DOI 10.1007/978-3-319-67934-1_13

are not up to the mark and some other algorithms fail to detect the ROI in a dynamic environment. However, they work fine in controlled environment.

For vehicle license plate detection - image acquisition, pre-processing, license plate localization and extraction of the license plate are the four basic steps. For license plate localization and extraction we are focusing on image segmentation and morphological image processing techniques respectively. As we know that localization and extraction of the license plate are the most important stages of VLP detection. Therefore we have proposed an improved image segmentation and license plate localization technique, based on Sliding Concentric window and morphological image processing. Where the SCW algorithm traverses the entire image and changes the value of each and every pixel to either 0 or 1, based on a comparison between a threshold value and the ratio of the statistical measurements of both of the windows [4, 5]. As a result, this algorithm keeps the pixels that have the possibility to be a part of the ROI. Additionally, the morphological image processing techniques named erosion and dilation uses structuring elements to recognize an objects shape with in an image [6]. Thus these techniques have been taken under consideration for the localization process of the license plate. We have implemented the SCW algorithm and the morphological approach separately in order to detect the license plate. However, the accuracy rate of the SCW algorithm was very poor and the accuracy of the morphological technique was not up to the mark. It can be concluded that the combined algorithms shall end up in a high accuracy. Where the input of the morphological approach will be the output of the SCW technique and the accuracy that we obtained from this approach is good enough.

The rest of the paper has been organized as follows, Sect. 2 describes the literature review, Sect. 3 illustrates the implementation criteria of the proposed model, Sect. 4 shows the performance analysis of the proposed model and a comparative study with Modified BernsensAlgorithm of VLP detection and Sect. 5 contains the concluding remarks of this study.

2 Literature Review

In this faster changing life style of people, an advanced transportation system has become a part and parcel. As a result, the number of vehicles along with the traffic is on the rise. Therefore the concept of automatic vehicle license plate (VLP) detection system for traffic control has emerged. Vehicle license plate detection is basically the process of identifying a vehicle by its number plate. But the identification process largely depends on image processing techniques such as image segmentation [3]. Several algorithms have been developed so far for the segmentation process. In this study, an optimum method of VLP detection system based on image segmentation and Morphological Image Processing has been developed to detect the license plate of a vehicle.

The first method of VLP detection was based on features of boundaries [7, 8]. In that technique, an image was binarized and then processed by using Hough transformation, to detect lines. Some methods, based upon the combinations of edge detection and mathematical morphology [9–12] showed very good results. In those techniques, the local variance and gradient magnitude of an image were computed. They were mainly focused on the change of brightness in the number plates region, which is more frequent and remarkable than otherwise. Gray-scale-based processing techniques were suggested in the literature of number plate localization [13, 14]. The reason behind, some algorithms do not provide a high degree of accuracy in case of detecting the VLP as the color is not steady and the lighting condition changes very frequently, in an image that contains natural scene. Furthermore, the fuzzy logic method has been adapted to locate number plate. In [15, 16], the authors made some instinctive rules to define the number plate and developed some membership functions for the fuzzy sets such as "bright", "dark", "bright and dark sequence", "texture" and "yellowness" to get the horizontal and vertical locations of number plate, but these approaches are subtle to the number plates color and brightness and need lengthier processing time in comparison with the conventional color-based techniques. Therefore, in spite of attaining better results, they still carry the downsides of the color-based schemes. Modified Bernsen algorithm was used to remove the shadow from an image, after horizontal and vertical correction and passing the image through median filtering. It was based on the fact that number plate's location is the region which has maximum histogram value [17].

Moving on, sliding concentric window (SCW) algorithm was used to keep the pixels that have similar characteristics as the pixels of the license plates region, then after implementing proper binarization and connected component analysis technique, the vehicle number plates location was determined. Such techniques fail to segment out the vehicles license plate having a black background and white characters on it [4, 5].

All the techniques mentioned above have some or many limitations. However, most of them perform much better in controlled environment. In this paper, for the detection of vehicle license plate, an improvised image segmentation technique named Sliding Concentric Window (SCW) and Morphological image processing method has been considered. Experimental results show that the proposed model of VLP detection depicts good results on inconstant environment, different lighting conditions and images that contain thenatural scene.

3 Implementation of the Proposed Model

This section contains a detail description of the two main algorithms that have been used for the implementation of the proposed model of VLP detection. Firstly the Sliding Concentric Window (SCW) algorithm, which was used for segmenting the input image. Secondly, the Morphological Image Processing techniques named erosion and dilation which were applied for the localization of the license plate. The proposed algorithm of this study is based on a hypothesis, that the features of the license plates region are not similar to the local characteristics of an image. Therefore the local characteristics of such images have been analyzed and an abrupt change has been observed at the license plates region. Based on that, the SCW algorithm locates the regions that tend to have irregularities in the local characteristics and manifest the presence of license plate [4]. Additionally, erosion and dilation have been included to localize the license plate. Figure 1 below exhibits the block diagram of the proposed model of VLP detection.

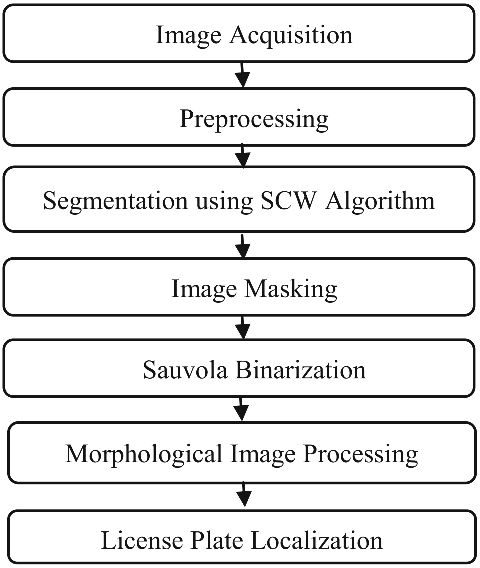


Image Acquisition
Preprocessing
Segmentation using SCW Algorithm
Image Masking
Sauvola Binarization
Morphological Image Processing
License Plate Localization

Fig. 1. Block diagram of the proposed model of vehicle license plate detection

3.1 Image Acquisition

It is the process of acquiring an image from some source for further processing. For this research, the images that have been chosen contains dynamic environment such as different lighting conditions, unwanted illuminations etc. The reason is real life scenario will not provide or work on the basis of an algorithms specification.

Therefore the algorithm should be compatible enough to detect the ROI from such circumstances. Some sample images from the dataset that has been chosen for the experiments of this research shown in Fig. 2.

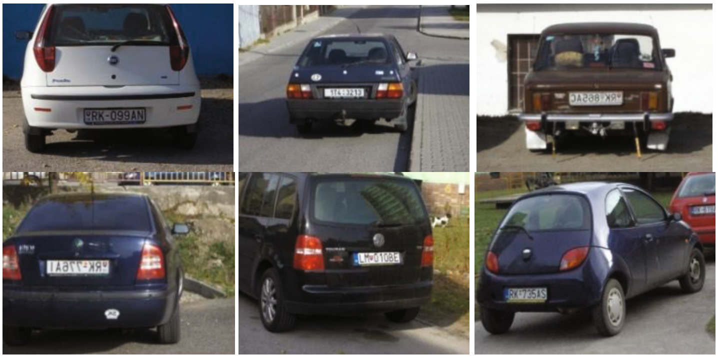


Fig. 2. Sample images from the dataset

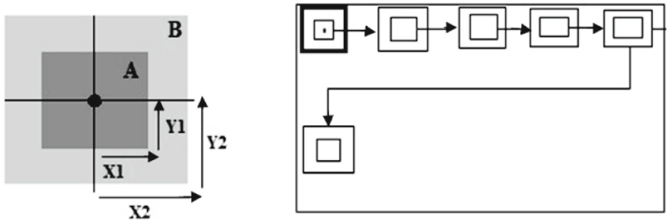
3.2 Preprocessing

In many cases, the step image preprocessing has been ignored, but it is one of the most important steps in the field of image processing [18]. Here in this study the input images have been preprocessed by using (i) Image resizing (ii) RGB to Gray scale conversion (iii) Median filtering for noise removal and (iv) Image inversion techniques, in order to get better accuracy rate in the segmentation stage. For resizing the image the aspect ratio of it has been preserved, by calculating the number of columns and by specifying the number of rows. In this case, the number of rows has been specified to 400. After that, the images have been converted to gray scale. The reason is the color of RGB image is not stable as a result the currently available solutions fail to provide a higher accuracy rate for the images that contain natural scene [5]. As one of the main purposes of this study is to detect the license plate from images that contain natural scene, therefore, the gray scale conversion process has a lot of significance. Furthermore, the median filtering method of noise removal has also been considered as one of the main parts of the pre-processing stage. Since the images are often get corrupted by noise. The reason is digital images are subjected to a wide variety of distortions during image acquisition [19]. Therefore recovering the original image from a noisy data is essential. Moving on to the next stages of preprocessing the input image has been inverted so that the accuracy rate of the SCW technique improves. As stated in [5], the SCW method does not guarantee to detect the ROI in the case of vehicles that have a dark background and white foreground or characters. Therefore the image inversion process has been conducted so that the SCW algorithm can perform the segmentation process accurately and can successfully detect the ROI from the input images [5]. The resulted image after the preprocessing stage has been shown in Fig. 4(b) and (c).

3.3 Segmentation Using SCW Algorithm

In order to establish the SCW algorithm the following steps have been carried out;

> **Step 1:** Two concentric windows A and B were created for some specific pixels of the input image. Where window B was two times bigger than window A in terms of height and width. In addition to that, the width of both of the windows was three times larger than their corresponding height and the center of them was the first pixel (Upper Left Corner) of the image, for the very beginning of the sliding process [4, 5]. For our research, the height and width for both of the windows A and B have been set to 2 and 6, 4 and 12 respectively. Here in Fig. 3(a) X1, Y1 and X2, Y2 are the height and width of the windows A and B respectively and the center of both of the windows (window A and B) was the first pixel of the image. This is the starting point of the sliding process of the two concentric windows. Both of the windows will slide till the end of the image (as shown in Fig. 3(b)) and the standard deviation or mean of the values of the pixels that are inside both of the windows will be calculated. At the same time, the ratio of the two windows statistical measurements have also been considered.
>
> **Step 2:** Through the above-mentioned process in Stage 1 each and every pixel of the input image will become the center of both of the windows only for once and



(a) (b)

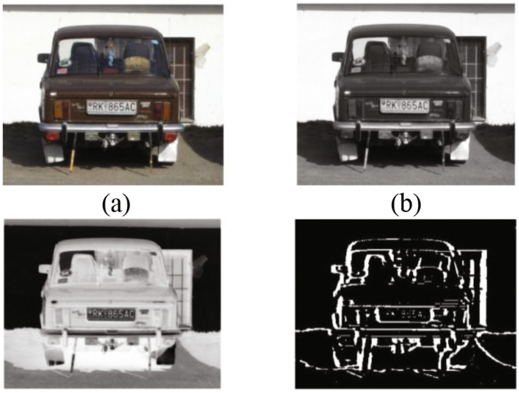
Fig. 3. (a) Concentric windows (b) windows scanning the image

eventually will be set to either 0 or 1 based on the comparison between both of the windows statistical measurements ratios and the threshold value.

If the ratio is less than the threshold value, which will be set by the user (in this case it's 1.02), then the center of the windows will be set to 0 otherwise 1. In the below equation (Eq. 1) *I1* is the gray image and *I2* is the resulted image after the sliding process which is basically a binary image.

$$\begin{aligned} &\text{in}\, I1(x,y)\; \mathit{if}\, \frac{\text{MA}}{\text{MB}} \le T\, \text{then}\; 12(x,y) = 0 \\ &\text{in}\, I1(x,y)\; \mathit{if}\, \frac{\text{MA}}{\text{MB}} > T\, \text{then}\; I2(x,y) = 1 \end{aligned} \tag{1}$$

The threshold value *T* has been determined by following the trial and error method [4, 5]. The reason is the algorithm checks the irregularities in the local characteristics of an image, therefore, the threshold value changes based on the environment of the



(a) (b) (c) (d)

Fig. 4. Steps of the SCW technique (a) input image, (b) *gray* image (c) image after the preprocessing stage and the inverted image, (d) resulted image after the completion of the sliding process and thresholding

dataset. In the above-illustrated equation (Eq. 1) *I1*(*x*, *y*) is the image in which the windows are sliding. *MA* and *MB* are the means or standard deviations of the pixels that are located at the inner and outer windows region respectively and *T* is the threshold value. At the end of this sliding process, we will get a binary image [4, 5]. The figure (Fig. 4(d)) above shows the resulted image after the SCW technique.

3.4 Image Masking

In this step, the intersection or logical AND operation, of the resulted image after the SCW process and the original image (Gray Image) has been calculated. Where the resulted of the SCW process has been considered as the mask [5]. It leads us to an image where we had the ROI along with some other extra information [4, 5]. In Fig. 5 (a), the resulted image after the image masking step has been given away. Where it is clear that the image masking stage presents an image where the vehicle edges can be seen clearly and the number plates region is also visible. Furthermore, to locate the ROI from the resulted image of the image masking phase, Sauvola Binarization technique has been conducted.

3.5 Sauvola Binarization Technique

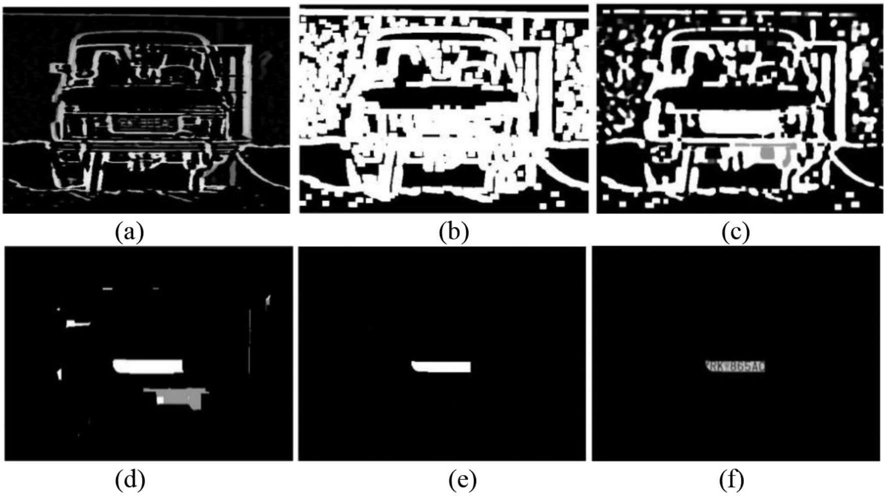
After the image masking process, the local adaptive thresholding method of Sauvola has been chosen for the binarization process of the output image. It performs a rapid classification of the local contents of the background of an image. The goal of the binarization algorithms is to produce an optimal threshold value for each pixel. The algorithm first computes a threshold for every n$^{\text{th}}$ pixel and then use interpolation for the rest of the pixels. Threshold computation is preceded by the selection of the proper binarization method based on an analysis of local image properties like range, variance and surface fitting parameters or their logical combinations. It is typical of locally adaptive methods to the coordinates x, y. This method adapts the threshold according to the local mean and standard deviation over a window size of b × b [20]. The threshold at pixel (x, y) is calculated as

$$T(x,y) = m(x,y).\left[1 + k.\left(\frac{s(x,y)}{R} - 1\right)\right] \tag{2}$$

Where *m*(*x*, *y*) and *s*(*x*, *y*) are the local sample mean and standard deviation respectively. Sauvola suggests the values of k = 0.5, R = 128 and b = 10, which were adopted in this algorithm. Thus the contribution of the standard deviation becomes adaptive. For example, for the badly illuminated areas, the threshold is lower [20]. Hence the image has been binarized by using the following equations (Eq. 3).

$$\left.\begin{array}{ll} I(x,y) = 1, & \mathit{if}\, I(x,y) \geq T(x,y) \\ I(x,y) = 0, & \mathit{if}\, I(x,y) < T(x,y) \end{array}\right\} \tag{3}$$

The figure below (Fig. 5(b)) illustrates the result after the Sauvola Binarization process.



(a) (b) (c)

(d) (e) (f)

Fig. 5. (a). Resulted in image after image masking process, (b) resulted image after sauvola binarization technique, (c) dilated image using disk structuring element, (d) eroded Image using rectangle structuring element, (e) eroded image using line structuring element and (f) localization of VLP

3.6 Morphological Image Processing

For the completion of the detection process of VLP, the resulted image after the Sauvola Binarization technique has been further processed by using Morphological Image processing. Where two of the most primary techniques have been adopted named as erosion and dilation. Dilation adds pixels to the boundaries and erosion removes pixels from the boundary of an object. The number of pixels added or removed from the objects in an image depends on the size and shape of the structuring element used to process the image [6].

Dilation: The morphological transformation dilation *"$\oplus$"* combines two sets using vector addition. The dilation $X \oplus B$ is the set of all possible vector additions of pairs of elements one from each of the sets X and B [6].

$$X \oplus B = \{\mathrm{p} \in \varepsilon 2\colon \mathrm{p} = \mathrm{x} + \mathrm{b},\ \mathrm{x} \in\ \mathrm{X\,and\,b}\ \in \mathrm{B}\} \tag{4}$$

In our work, we dilated the gray scale image to improve the given image by filling holes, objects edges sharpening and broken lines joining. Where the disk structuring element has been used. Figure (Fig. 5(c)) illustrates the output of the dilation process.

Erosion: In [6], Erosion ⊖ combines two sets using vector subtraction of set elements and is the dual operator of dilation. Neither erosion nor dilation is an invertible transformation

$$X \ominus B = \{\mathrm{p} \in \varepsilon 2 \colon \mathrm{p} = \mathrm{x} + \mathrm{b},\ \mathrm{x} \in \ \mathrm{X}\,\mathrm{for\,every\,b} \in \ \mathrm{B}\} \tag{5}$$

This formula (Eq. 5) says that every point p of the image was tested; the result of the erosion has been given by those points p for which all possible p + b was in X. Erosion has been used for thinning the edges of the binary image by using 20 × 20 rectangular shape structuring element [6] and after that line structuring element has been used for further erosion process. The resulted image after the completion of the erosion technique has been shown in Fig. 5(d) and (e).

3.7 License Plate Localization

In order to localize the license plate the image masking step has been taken under consideration again. It was the same process that has been conducted after the SCW technique. The only difference is the mask that has been used, in this case it was the resulted image after the completion of the Morphological Image Processing technique. Figure (Fig. 5(e)) shows the mask that was used for the license plate localization process. Figure 5(f) exhibits the resulted image after the license plate localization method and it can be clearly seen that our proposed model of vehicle license plate detection can successfully detect the license plate of a vehicle.

4 Result Analysis

All the experiments of this research were conducted by using MATLAB development environment. We applied our proposed model of VLP detection over 97 images, the images were captured from various lighting conditions and natural backgrounds, and the algorithm was able to successfully detect 84 license plates from the selected dataset. The accuracy rate of the proposed model of VLP detection algorithm is 86.5%. In Table, I, below the variation of the accuracy rate of the proposed algorithm, based on the changes in the threshold value, has been given.

As the threshold is determined by using a trial and error procedure, therefore, it varies from dataset to dataset [4, 5]. In the below graph (Fig. 6) it can be noticed that the accuracy is at the peak when the threshold is 1.02. At first, the accuracy rate was increasing but then after one certain point (threshold 1.02 and accuracy 86.50%), it has started to decrease. However, at 1.06, the accuracy rate has slightly increased in comparison with the other thresholds but it was less than the accuracy rate of the threshold 1.02. Therefore 1.02 has been considered as the optimum threshold value for the experimented dataset (Table 1).

Another important determinant of the accuracy rate is the window size of the SCW process. The perfection of the detection process of the algorithm varies from window size to window size. Therefore selecting the optimum window size has a lot of significance. As we have seen that the accuracy rate is high when the threshold is 1.02,

Table 1. Accuracy rate comparison on different threshold value

Threshold value	Accuracy rate (%)
0.99	28.87
1.00	48.45
1.01	75.26
1.02	86.50
1.03	83.50
1.04	80.42
1.05	71.13
1.06	76.29
1.07	67.01
1.08	56.70
1.09	50.51
1.10	45.36

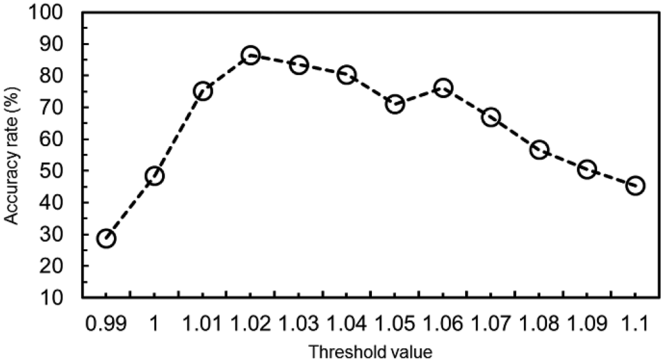


Fig. 6. Variation of accuracy rate as the threshold value increases

therefore, we have kept the threshold same and changed the window size of the SCW process and observed the changes in the accuracy rate of different window sizes. Table 2 shows the variation of the accuracy rate based on the changes in the window size.

Table 2 illustrates that the optimum size for Window "A" is Height = 2 and Width = 6 and Window "B" is Height = 4 and Width = 12 because the number of detected license plates was 84. We can also see that the number of detected license plates for this window size was significantly higher in comparison with the others. As a result, it has been considered as the most suitable window size for our dataset.

Moving on, in this section, a comparison between the Improved Bernsen algorithm [17] (histogram based VLP detection) and the proposed model of VLP detection process has been illustrated. As it is one of the major purposes of this research. We have run both of the algorithms on the same dataset and observed that the BernsensAlgorithm fails to detect the license plate from images that contain a natural background. Although it works well when the images were focused on the license plates region. But if the input

Table 2. Variation of the accuracy rate based on different window size

Window A [height, width]	Window B [height, width]	License plate detected (out of 97)
[1, 3]	[2, 6]	42
[2, 6]	[4, 12]	84
[3, 9]	[6, 18]	32
[4, 12]	[8, 24]	22
[5, 15]	[10, 30]	17
[6, 18]	[12, 36]	16
[7, 21]	[14, 42]	16

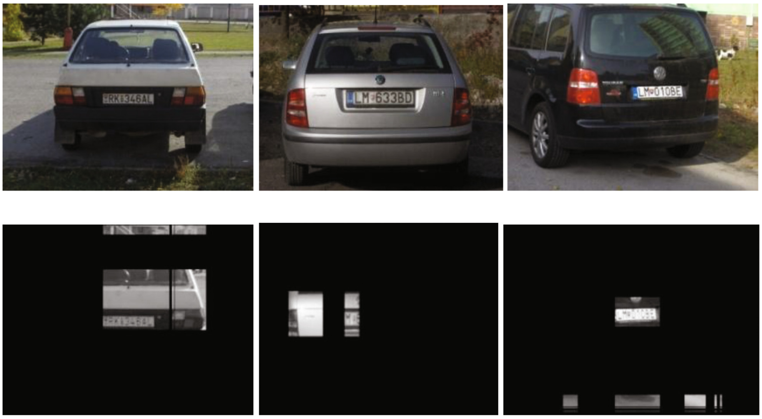


Fig. 7. Brensen algorithms failure in case of VLP detection

image contains illumination then the algorithm provides an output where the resulted image contains multiple segments of the image as the ROI. In some other cases, it detects some other portion of the input image as the ROI which is not the license plate (Fig. 7).

Table 3 represents the accuracy rate of both of the algorithms. The simulation has been conducted over the same dataset that consists of 97 images. The proposed algorithm has successfully detected 84 license plates where the Bernsens algorithm detected 57 license plates and provides an accuracy rate of 58.76%.

Table 3. Comparison between the Proposed and Bernsen Algorithms

Algorithm	Accuracy rate (%)
Proposed model	86.50
Improved bernsens algorithm	58.76

5 Conclusions

In this study, the entire process of VLP detection was divided into two sections. The first one was image segmentation where Sliding Concentric Window algorithm has been implemented to segment out the ROI. The other section was based on Morphological Image processing where the resulted image after the segmentation process has been dilated and then eroded twice by using the disk, rectangle and line structuring element respectively. The experimental results show that the proposed model VLP detection was capable enough to successfully detect the license plate from images that contain natural scene and different lighting conditions. Moreover, the proposed model is robust enough to detect the ROI from images where the angle of the captured image is also different. As it has been noticed that the accuracy rate is high enough, therefore, this algorithm can be considered as an optimum algorithm for vehicle license plate detection.

Acknowledgements. We would like to acknowledge BRAC University, Bangladesh to provide their facilities and Dr Md. Moinul Hossain for the useful discussions. A very special acknowledgement is also made to the referees who make important comments to improve this paper.

References

1. Tian, B., Yao, Q., Gu, Y., Wang, K., Li, Y.: Video processing techniques for traffic flow monitoring: a survey. In: 14th International IEEE Conference on Intelligent Transportation Systems (ITSC), pp. 1103–1108 (2011)
2. Barcellos, P., Bouvie, C., Escouto, F.L., Scharcanski, J.: A novel video based system for detecting and counting vehicles at user-defined virtual loops. Expert Syst. Appl. **42**, 1845–1856 (2015)
3. Tan, X.-J., JunLiu, C.: A video-based real-time vehicle detection method by classified background learning. World Trans. Eng. Technol. Edu. **6**, 189 (2007)
4. Anagnostopoulos, C., Anagnostopoulos, I., Tsekouras, G., Kouzas, G., Loumos, V., Kayafas, E.: Using sliding concentric windows for license plate segmentation and processing. In: IEEE Workshop on Signal Processing Systems Design and Implementation, pp. 337–342, November 2005
5. Anagnostopoulos, C., Anagnostopoulos, I., Loumos, V., Kayafas, E.: A license plate-recognition algorithm for intelligent transportation system applications. IEEE Trans. Intell. Transp. Syst. **7**(3), 377–392 (2006)
6. Sonka, M., Vaclav H., Boyle, R.D.: Mathematical morphology. Image processing, analysis, and machine vision. In: International Student edn. Thompson Learning, Toronto, pp. 657–664 (2008)
7. Kamat, V., Ganesan, S.: An efficient implementation of the hough transform for detecting vehicle license plates using DSP's. In: Proceedings of Real Time Technology and Applications, Chicago, 15-17 May 1995, pp. 58–59 (1995)
8. Yanamura, Y., Goto, M., Nishiyama, D.: Extraction and tracking of the license plate using hough transform and voted block matching. IEEE proceedings of Intelligent Vehicles Symposium, Columbus, 9-11 June 2003, pp. 243–246 (2003)
9. Martín, F., García, M., Alba, L.: New methods for automatic reading of VLP's (Vehicle License Plates). In: Proceeding of IASTED International Conference SPPRA, June 2002

10. Hongliangand, B., Changping, L.: A hybrid license plate extraction method based on edge statistics and morphology. In: Proceeding of ICPR, pp. 831–834 (2004)
11. Zheng, D., Zhao, Y., Wang, J.: An efficient method of license plate location. Pattern Recognit. Lett. **26**(15), 2431–2438 (2005)
12. Lee, H.-J., Chen, S.-Y., Wang, S.-Z.: Extraction and recognition of license plates of motorcycles and vehicles on highways. In: Proceeding of ICPR, pp. 356–359 (2004)
13. Shi, X., Zhao, W., Shen, Y.: Automatic license plate recognition system based on color image processing. In: Gervasi, O., et al. (eds.) Computational Science and Its Applications – ICCSA 2005. ICCSA 2005. Lecture Notes in Computer Science, vol 3483, pp. 1159–1168. Springer, New York(2005)
14. Yan, D., Hongqing, M., Jilin, L., Langang, L.: A high performance license plate recognition system based on the web technique. In: Proceeding of Conference Intelligent Transportation Systems, pp. 325–329 (2001)
15. Zimic, N.,. Ficzko, J., Mraz, M., Virant, J.: The fuzzy logic approach to the car numberplate locating problem. In: Proceeding IIS, pp. 227–230 (1997)
16. Chang, S.-L., Chen, L.-S., Chung, Y.-C., Chen, S.-W.: Automatic license plate recognition. IEEE Trans. Intell. Transp. Syst. **5**(1), 42–53 (2004)
17. Latha, M.G., Chakravarthy, G.: An improved Bernsen algorithm approaches for license plate recognition. IOSR-JECE: IOSR J. Electron. Commun. Eng. **3**(4), 01–05 (2012)
18. Wang, T.-H., Ni, F.-C., Li, K.-T., Chen, Y.-P.: Robust license plate recognition based on dynamic projection warping. In: Proceeding IEEE International Conference Networking, Sensing and Control, pp. 784–788 (2004)
19. Rong, Z., Yong, W.: Application of improved median filter on image processing. J. Comput. **7**(4), 838–841 (2012)
20. Sauvola, J., Pietikäinen, M.: Adaptive document image binarization. Pattern Recogn. **33**, 225–236 (2000)

An Improved Approach for Securing Document Images Using Dual Cover

R.E. Vinodhini(✉), P. Malathi, and T. Gireesh Kumar

Department of Computer Science and Engineering,
Amrita School of Engineering, Coimbatore,
Amrita Vishwa Vidyapeetham, Amrita University, Coimbatore, India
cb.en.p2cse15029@cb.students.amrita.edu,
{p_malathy, t_gireeshkumar}@cb.amrita.edu

Abstract. Security is essential to all varieties of data that is transmitted through an open network or the internet. Steganography is such a technique which provides security for any kind of data by hiding it inside a cover object. The images with confidential information like medical reports, passport, academic certificates, Aadhar card and agreements between governments need to be secured while transferring through the internet. In this paper, dual cover steganography is used to provide a better security which prevents the confidential data from attackers. This approach uses two layers of covers, image, and DNA. Using improved DNA insertion method the document image is hidden inside a DNA sequence. The BPN, capacity, payload and embedding time are calculated for the DNA sequence. The improved insertion method is used in this paper since it gives the very low cracking probability. The fake DNA is hidden inside a cover image using matrix embedding with hamming code algorithm for both spatial and transform domain of an image. The Mean Square Error (MSE), Peak Signal Noise Ratio (PSNR), Maximum difference, average difference, structural content are calculated and compared. To ensure the security of document image the statistical attack called histogram analysis attack is performed and compared. This comparison shows that the matrix embedding with hamming code in the frequency domain provides better security for the second layer of hiding the data.

Keywords: Dual cover · DNA · DNA hiding · Hamming codes · DWT

1 Introduction

Steganography is the classical technique of securing the top secret information of a person, organization, and government, etc. [1]. The cover object is concealed so that the data hidden inside the cover object is not visible to the human eyes. There are a lot of steganography algorithms based on the cover objects used, provides security at different levels. Image steganography is one of the most used steganography technique that uses the image as a cover medium for secret data hiding. The secret data size that can be hidden inside the cover image is determined based on the embedding capacity of the image. The embedding capacity of an image is the number of bits that can be replaced to hide a secret data [2]. The DNA steganography is a secured form of

S.M. Thampi et al. (eds.), *Advances in Signal Processing and Intelligent Recognition Systems*, Advances in Intelligent Systems and Computing 678,
DOI 10.1007/978-3-319-67934-1_14

steganography, which helps to hide the secret data in the DNA sequence. The DNA sequence is the theoretical form of human DNA, which consists of more than ten million sequences. The theoretical of the DNA has four bases like Adenine, Thymine, Cytosine, and Guanine [3]. The data of higher length can be hidden inside the DNA sequence since it has a large number bases in sequences. The hidden data is practically impossible to find by an attacker using the existing steganalysis techniques. Though it provides high security to the secret data than image steganography, it has its own drawback too [4]. To overcome this disadvantage a dual cover steganography is introduced, that uses DNA and image as two layers of the cover object. This provides a better security than image and DNA steganography. The data that have to be transferred through the internet or an open network needs to be secured in order to prevent the confidential information from the attackers or unauthorized users. The data can be of secret information like username and password of various accounts (bank, institution, and company accounts), ATM card information, passport details, medical reports, the agreement between governments and academic certificates. The data can be in many forms like alphabets, alphanumeric, images, audios, and videos.

In this paper, a dual cover steganography approach is used to hide a document image inside dual covers of imaging and DNA. The DNA steganography technique improved insertion method is used to hide the document image inside the human DNA. The fake DNA is obtained with the document image hidden inside the original DNA. The BPN, capacity, payload and embedding time are calculated for the DNA sequence [5]. The cracking probability of the improved insertion technique is very low and practically it is impossible to crack by an attacker. Since the exchange of DNA sequence through internet leads to the loss of some DNA bases and thus the whole meaning of the hidden message may be affected. To avoid this fake DNA is hidden in the cover image for ease of transmission. The image steganography techniques like matrix embedding with Hamming code algorithm for both spatial and frequency domain of an image are used in this paper. The comparisons of these algorithms are done by MSE, PSNR, Maximum difference, Average difference, Structural content. The confidentiality of the hidden document image is tested using a statistical attack called histogram analysis attack.

The rest of this paper is organized as follows: Sect. 2 explains related works, problem formulation is done in Sects. 3 and 4 states the performance metrics, Sect. 5 lists steganalysis methods, Sect. 6 gives the brief of results and discussions, Sect. 7 summarizes the implementation results and the conclusion with future enhancement of the research work is done in Sect. 8.

2 Related Works

Della et al. [6] proposed a novel technique based on DWT for securing image using steganography. In this paper, a data hiding technique is used for hiding multiple colour images into a single colour image using DWT. The RGB planes are split from the cover image and the secret message is embedded into these planes. The PSNR, SSIM values are calculated for ensuring the quality of the proposed work.

Malathi et al. [7] proposed a paper that relates the embedding efficiency of the LSB technique in the domain of spatial and transform in image steganography. With this approach, F5 and matrix embedding are combined with the least significant bit method. The outcome of the techniques are compared by using histograms and the performance is measured by using the PSNR and MSE values.

Wang et al. [8] proposed a paper on information hiding based on DNA steganography. This paper uses cryptography scheme called vigenere ciphers to increase the security of the secret information. The cipher text obtained from the vigenere cipher is coded and then it is converted into binary code. The using the DNA encoding scheme the message is hidden inside the DNA sequence. This paper also takes care of the transmission part of the secret information.

Menaka [9] proposed a paper on the encryption of messages using DNA sequences. The dictionary method is used for the conversion of DNA base into binary codes. There are three complement rules used in this paper which help to hide the message in the DNA sequence. The encoded DNA sequence is transmitted over the open network to the recipient and the receiver decodes the fake DNA sequence using the inverse of the complementary rule used for the encoding process.

Peterson et al. [10] developed a system which hides the secret data in the reference DNA sequence by substituting the DNA bases with the message bits. There are totally 64 symbols used in this paper for the purpose of encoding. For example D = TTG, S = ACG and so on. The letters E and I most frequently used words in the English alphabets and this helps the attackers to crack the hidden message. The data hiding in the DNA uses the biological properties and it is not economically efficient to implement.

For the purpose of dual cover steganography Das et al. [11] developed a system by using a chaotic map in DNA. The overall security enhancement is done on the existing steganography techniques. The dual covers like image and DNA are used in this system to improve the security. In this method, the two bits of secret data are hidden in a single pixel. The capacity check is performed based on the length of hidden message. The fake DNA is hidden in the cover image with the help of the generated data code replacement. The extraction process reverses the embedding process and it also uses the same KEY3 used for embedding.

3 Problem Formulation

The proposed work uses image and DNA as dual covers in order to overcome the limitations of hiding the secret data inside an image or a DNA sequence.

The document images like medical reports, passport, Aadhar card, academic certificates and agreements between governments are used in this approach. The document image may be of different sizes based on the quality and content of the document. The document image is the input to the system which is covered inside dual cover objects. The input is resized till 512 × 512 and converted into a binary form using the binary encoding of the image. The DNA sequence is the layer1 cover object which takes the sequence of DNA bases based on the length of the binary sequence of the input. The DNA sequence is converted into a binary using dictionary method [12] and by using the key value it is segmented. The key value is a randomly generated value and it is kept

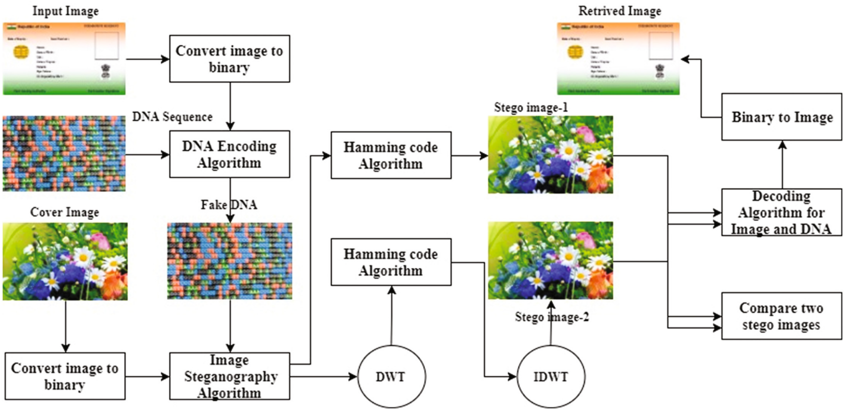


Fig. 1. Proposed architecture

low to have the DNA sequence of minimum length to hide the document image. In this scheme, the key value is fixed to 3 and the binary converted DNA is segmented by using the key value. The key value is binary converted and XOR-ed with the binary value of the document image. Using DNA improved insertion algorithm the XOR-ed results are placed one by one in the most significant bit of the DNA segments. The segments are then concatenated to form a sequence of binary bits. Using Dictionary method the DNA bases are converted from the binary sequenced and its forms the fake DNA sequence. The BPN, capacity, payload and embedding time are calculated from the DNA sequence. The cracking probability of the proposed technique is less than 0.00001 and practically it is impossible to crack by an attacker. This fake DNA sequence is not safe to transfer to other systems due to the chance of losing some DNA bases. To avoid that the fake DNA sequence is hidden in the cover image. In this paper, the fake DNA sequence is converted into a binary sequence using the dictionary method. The RGB image that acts as a cover object is binary converted and techniques like matrix embedding using hamming code algorithm [13] are applied in both spatial and transform domain to hide the fake DNA into the image. The MSE, PSNR, AD, MD, SC values are calculated to compare the quality of the stego-image on various techniques used. To ensure the security of confidentiality of the hidden document image, the statistical attack called histogram analysis attack is performed and compared.

To retrieve the document image the process is reversed. The stego-image is binary converted by using the inverse of the used image steganography technique the binary sequence of the fake DNA is retrieved. Then the Fake DNA is segmented using the key value (k = (previous K) + 1) and from the segments, the most significant bits are separated and concatenated to form a binary sequence of the document image. This binary sequence is converted into an image, which is a document image. The whole process of the proposed work is shown in Fig. 1.

3.1 Proposed DNA Algorithm

3.1.1 DNA Embedding Algorithm

//Input: Secret data (Document Image) S € {0, 1} ^{S}, Reference DNA d € {0, 1} ^{s}
//Output: Fake DNA sequence with hidden document image

Step 1: The document image S is converted into binary sequence.
Step 2: The reference DNA D is binary converted by using the dictionary method (A = 00, C = 01, G = 10, T = 11).
Step 3: The random sequence R_1, R_2, R_n are generated and the value t is found using the formula $E_i=\sum_{n=1}^{i} R_n > |S|$ and the key value is chosen from the random sequence R_1, R_2, R_{n-1}.
Step 4: Choose key value from the above random sequence R_1, R_2, R_{t-1}. (Here k=3) and equally segment D with k value results d_1, d_2
Step 5: The binary value of the chosen key k is XOR-ed with S.
Step 6: The binary sequence of S_t in appended to front d_1, d_2.
Step 7: Transform the binary sequence to DNA base using dictionary method.

3.1.2 DNA Decoding Algorithm

//Input: Fake DNA f € {0, 1} ^{s}
//Output: secret data (Document Image), Reference DNA

Step 1: The fake DNA f is binary converted by using the dictionary method (A = 00, C = 01, G = 10, T = 11)
Step 2: Using random seed U and V, generate random sequence U_1, U_2….U_n…and V_1, V_2…V_n…
Step 3: The value p is found using the formula $\sum_{n=1}^{t}(U_n + V_n) \leq |f_1|$and the sequence formed is U_1+V_1…U_P+V_P, then the fake DNA f is segmented using the sequence.
Step 4: For each segment n, 1≤n, ≤p, extract first U_n bits called $S_{n.}$ For segment n, 1≤n≤p, extract last $V_n$bits called f_n.
Step 5: Concatenate all f_n, 1≤n≤,p, results the reference DNA by applying inverse function of dictionary method.
Step 6: Concatenate all S_n, 1≤n≤p, and the Binary value of S_n and (n-1) are XORed which results the document image.

3.2 Matrix Embedding Using Binary Hamming Code

The matrix embedding helps to increase the embedding capacity of the cover object by flipping the LSB bits or by using the +/− embedding method. The number of steps performed in the matrix embedding is high and difficult than the other techniques, so it provides better security than the other methods [14]. The matrix embedding algorithm using binary hamming code is used in this paper to increase the embedding capacity of the cover object and also to give better security than the other methods [15]. The steps for the matrix embedding with hamming code are given in the form of a flow diagram (Fig. 2).

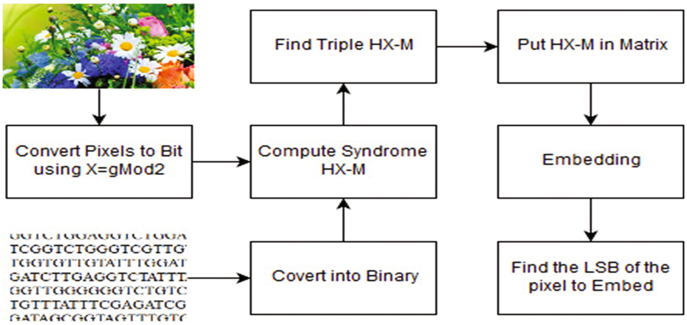


Fig. 2. Flow diagram for hamming code

3.2.1 Proposed Image Algorithm

The transform domain tools group has algorithms like Discrete Wavelet Transform (DWT), Fast Fourier Transform (FFT), and Discrete Cosine Transform (DCT). The DWT algorithms are based on wavelets which are used for the compression and transmission also analysis of the images. The DWT has great characteristics in time and frequency domain. Comparing to the DCT the DWT provides reliable coding efficiency and provide a better quality of restored images. The LL – Horizontal and vertical low pass LH – Horizontal low pass and vertical high pass HL – Horizontal high pass and vertical low pass HH – Horizontal and vertical high pass are four components of DWT [16] (Fig. 3).

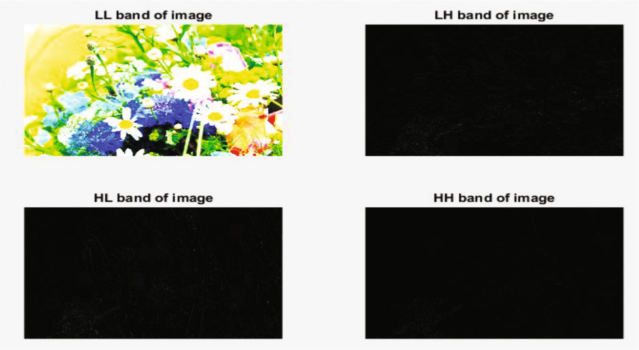


Fig. 3. DWT applied image

The LL band is the only color plane of the four existing bands. The LH, HL, HH are the bands which is used to embed the secret message (Fig. 4).

The above image is the work flow of applying DWT with Hamming code algorithm for hiding the fake DNA inside the image. The cover images split into four bands of LL, LH, HL, and HH by applying DWT algorithm. The LH, HH, HL bands of the cover image are used for the purpose of embedding the fake DNA inside the cover image. The hamming code algorithm is used here for the embedding process.

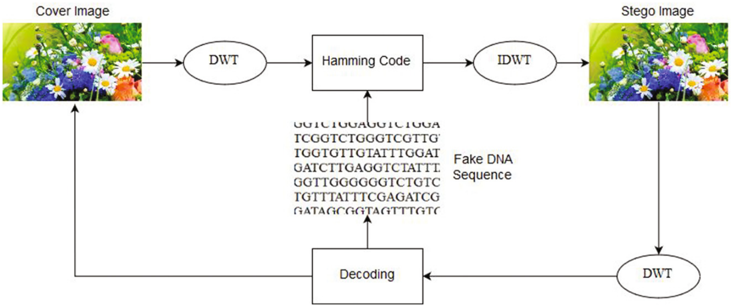


Fig. 4. Flow diagram for transform domain

The decoding is done to retrieve the actual DNA sequence. The inverse DWT is performed to restore the original cover image.

4 Performance Metrics

The BPN, Capacity, Payload and Cracking probability are calculated for measuring the quality of the fake DNA. The metrics like Peak signal to noise ratio (PSNR), Mean squared error (MSE), Average difference (AD), Maximum difference (MD) and Structural content (SC) are calculated for confirming the quality of the stego image [17]. The cracking probability is calculated by the following formula

$$\mathbf{P(r)} = \frac{\mathbf{1}}{\mathbf{1.63 \times 10^8}} \times \frac{\mathbf{1}}{\mathbf{n-1}} \times \frac{\mathbf{1}}{\mathbf{2^m-1}} \times \frac{\mathbf{1}}{\mathbf{2^s-1}} \times \frac{\mathbf{1}}{\mathbf{24}} \times \frac{\mathbf{1}}{\mathbf{2^{3m}}}, \text{where } n,\ m,\ s \geq 1 \quad (1)$$

Here,

- n is the number of bits in the Fake DNA sequence
- m is the number of bits in the secret data
- s is the number of bits in the reference DNA sequence

Here,

$\frac{1}{1.63 \times 10^8}$ Gives the probability to find the DNA sequence. Hence, 1.63×10^8 DNA available in internet. $\frac{1}{n-1}$ Is the probability to predict the number of fake DNA sequence, $\frac{1}{2^m-1}$ Gives the probability to predict the length of the secret data, $\frac{1}{2^s-1}$ Gives the probability to predict the number of bits in the reference DNA sequence, $\frac{1}{24}$ Gives the probability to find the Dictionary coding rule used and $\frac{1}{2^{3m}}$ Gives the probability to find the XOR rule (Table 1).

The proposed approach for the DNA hiding part provides low cracking probability than the existing algorithms.

Table 1. Cracking probability of existing and proposed algorithm

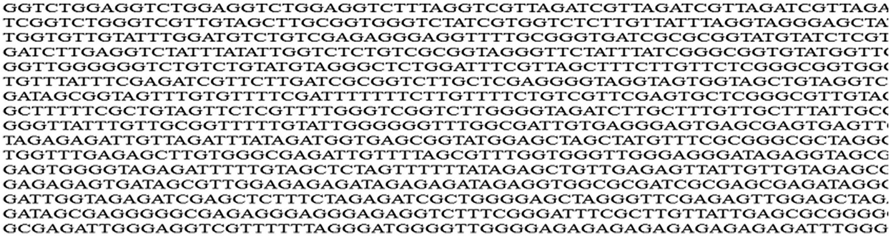
Methods	Cracking probability
Insertion method	$\frac{1}{1.63*10^8} \times \frac{1}{24} \times \frac{1}{(n-1)} \times \frac{1}{(2^m-1)} \times \frac{1}{2^{s-1}}$
Complementary method	$\frac{1}{1.63*10^8} \times \frac{1}{24^2}$
Substitution method	$\frac{1}{1.63*10^8} \times \frac{1}{3^n}$
Proposed method	$\frac{1}{1.63*10^8} \times \frac{1}{24} \times \frac{1}{(n-1)} \times \frac{1}{(2^m-1)} \times \frac{1}{2^{s-1}} \times \frac{1}{2^{3m}}$

5 Results and Discussion

The implementation of the system is carried out using MATLAB 2015 and the results are compared. The size of the document image used as the cover images are 32 × 32, 64 × 64, 128 × 128, 256 × 256, 400 × 400 and 512 × 512. The techniques like DNA improved insertion, matrix embedding using hamming codes are performed in the spatial and transform domain of the cover image.

5.1 Using the First Layer Cover Object "DNA"

The document image is binary converted and inserted into the DNA sequence using the DNA encoding algorithm. The fake is produced by having the document image inside it and it is shown in the below figure (Fig. 5).



```
GGTCTGGAGGTCTGGAGGTCTGGAGGTCTTTAGGTCGTTAGATCGTTAGATCGTTAGATCGTTAGA
TCGGTCTGGGTCGTTGTAGCTTGCGGTGGGTCTATCGTGGTCTCTTGTTATTTAGGTAGGGAGCTA
TGGTGTTGTATTTGGATGTCTGTCGAGAGGGAGGTTTTGCGGGTGATCGCGCGGTATGTATCTCGT
GATCTTGAGGTCTATTTATATTGGTCTCTGTCGCGGTAGGGTTCTATTTATCGGGCGGTGTATGGTT
GGTTGGGGGGTCTGTCTGTATGTAGGGCTCTGGATTTCGTTAGCTTTCTTGTTCTCGGGCGGTGG
TGTTTATTTCGAGATCGTTCTTGATCGCGGTCTTGCTCGAGGGGTAGGTAGTGGTAGCTGTAGGTC
GATAGCGGTAGTTTGTGTTTTCGATTTTTTTCTTGTTTTCTGTCGTTCGAGTGCTCGGGCGTTGTA
GCTTTTTCGCTGTAGTTCTCGTTTTGGGTCGGTCTTGGGGTAGATCTTGCTTTGTTGCTTTATTGC
GGGTTATTTGTTGCGGTTTTTGTATTGGGGGGTTTGGCGATTGTGAGGGAGTGAGCGAGTGAGTT
TAGAGAGATTGTTAGATTTATAGATGGTGAGCGGTATGGAGCTAGCTATGTTTCGCGGGCGCTAGG
TGGTTTGAGAGCTTGTGGGCGAGATTGTTTTAGCGTTTGGTGGGTTGGGAGGGATAGAGGTAGCG
GAGTGGGGTAGAGATTTTTGTAGCTCTAGTTTTTTATAGAGCTGTTGAGAGTTATTGTTGTAGAGCC
GAGAGAGTGATAGCGTTGGAGAGAGATAGAGAGATAGAGGTGGCGCGATCGCGAGCGAGATAGG
GATTGGTAGAGATCGAGCTCTTTCTAGAGATCGCTGGGGAGCTAGGGTTCGAGAGTTGGAGCTAG
GATAGCGAGGGGGGCGAGAGGGAGGGAGAGGTCTTTCGGGATTTCGCTTGTTATTGAGCGCGGGG
GCGAGATTGGGAGGTCGTTTTTTAGGGATGGGGTTGGGGAGAGAGAGAGAGAGAGATTTGGG
```

Fig. 5. Fake DNA

The document image is taken in the sizes from 32 × 32 to 512 × 512 in order to calculate the payload, capacity, BPN and embedding time for each size of the document image. Here payload is the size of new sequence after extracting out the reference DNA sequence. Capacity is the total length of the fake DNA sequence. BPN (Bits per Node) is number of bits hidden per characters. Embedding time is the time taken to hide the secret data in the DNA (Table 2).

Table 2. Result and analysis using the cover "DNA"

Document image size	Payload	Capacity	BPN	Embedding time
32 × 32	128	16512	0.022	0.2028 s
64 × 64	256	65792	0.020	0.4524 s
128 × 128	512	262696	0.018	2.2621 s
256 × 256	1024	1049600	0.004	7.7376 s
400 × 400	1600	2561600	0.007	16.9100 s
512 × 512	2048	4196152	0.004	27.2845 s

5.2 Using the First Layer Cover Object "DNA"

5.2.1 Image Results (Spatial and Transform Domain)

The size of the cover image is 1200 × 1200 pixels and it is a high-quality RGB image. The cover image and the stego image is of the same size with the secret data embedded within that image (Fig. 6).

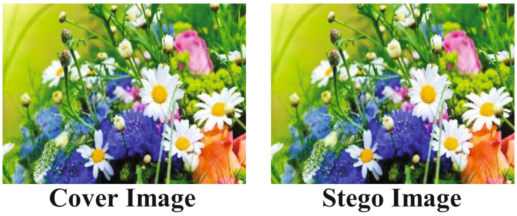

Cover Image
Stego Image

Fig. 6. Cover and stego images of hamming code technique (1200 × 1200) for spatial domain and transform domain

5.2.2 Results for Spatial Domain in Image (Existing Work)

See Table 3

Table 3. Result and analysis using the hamming code in spatial domain

S. No.	Length of the fake DNA(bits)	MSE (error)	PSNR (dB)	MD	AD	SC
1	131072	0.117985	57.412	3.5	−0.03	0.999
2	524288	0.153380	56.273	3.5	−0.05	0.999
3	2097152	0.160714	56.070	3.6	−0.130	0.998
4	8388608	0.173788	55.730	3.6	−0.136	0.9988
5	20480000	0.199617	55.128	3.6	−0.299	0.9988
6	33554432	0.203763	55.039	3.6	−0.3	0.999

5.2.2.1 Steganalysis for Spatial Domain in Image(Existing Work)

Histogram attack, the first histogram represents the histogram of the cover image and the second represents of stego image. Here, the histograms of the cover and stego

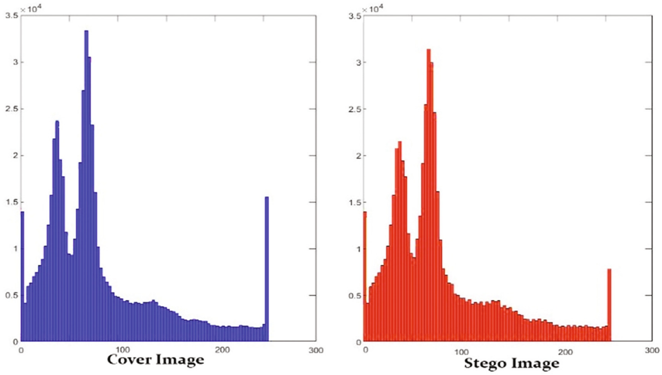


Fig. 7. Histogram for hamming code technique (spatial)

images have small variation and using this histogram it is possible to find there is some data is hidden (Fig. 7).

5.2.3 Results for Transform Domain in Image(Proposed Work)

See Table 4

Table 4. Result and analysis using hamming code in transform domain

S. No.	Length of the fake DNA(bits)	MSE (error)	PSNR (dB)	MD	AD	SC
1	131072	0.025510	64.063	2.0	0.00007	1.000004
2	524288	0.033801	62.841	2.0	0.00023	1.000007
3	2097152	0.039860	62.125	2.0	0.00057	1.000002
4	8388608	0.047832	61.333	2.0	0.00009	1.000011
5	20480000	0.071429	59.592	2.0	0.00065	1.000014
6	33554432	0.103316	57.989	2.0	0.00089	1.000009

5.2.2.2 Steganalysis for Transform Domain

Histogram attack, the first histogram represents the histogram of the cover image and the second represents of stego image.

Here, both the histograms of the cover and stego images are similar and the variation is not visible to the human eye so that using histogram attack it is not possible to find whether find any data is hidden or not (Fig. 8).

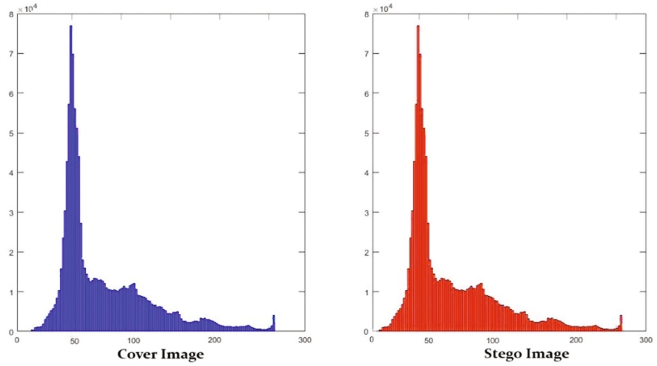


Fig. 8. Histogram for hamming code technique (transform)

6 Implementation and Evaluation

The above chart represents the comparison between PSNR values of Matrix embedding using hamming code algorithm in spatial and transform domain of an image (Fig. 9).

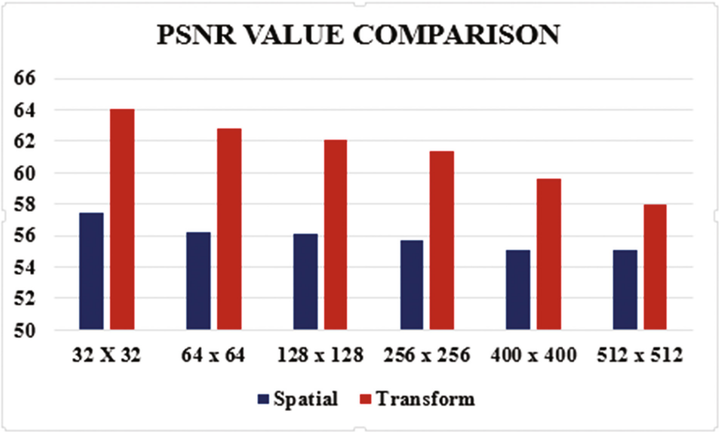


Fig. 9. PSNR value comparison chart

7 Conclusion

The designed system steganography with dual covers provides more security than the other existing techniques. The DNA sequence and RGB Image are taken as dual covers to hide the document image in-order to provide higher security for it. The document image is hidden in secondary cover object DNA sequence by using an improved insertion method. The obtained fake DNA is hidden inside the primary cover object RGB image by using image steganography techniques like matrix embedding using hamming code algorithms in spatial (Existing) and transform domain (proposed) of an image. The MSE, PSNR, MD, AC, SD values are calculated for the spatial (Existing) and transform domain (Proposed) images and they are compared. The statistical attack is performed on the proposed system and the cracking probability is calculated for the proposed algorithm. The improved insertion technique has the lowest cracking probability in DNA steganography and matrix embedding using a Hamming code for transform domain used in this work outperforms the existing image steganography algorithms in the spatial domain to hide the secret message.

References

1. Shih, F.Y.: Digital Watermarking and Steganography: Fundamentals and Techniques. CRC Press, Boca Raton (2017)
2. Subhedar, M.S., Mankar, V.H.: Current status and key issues in image steganography: a survey. Comput. sci.rev. **13**, 95–113 (2014)
3. Marwan, S., Shawish, A., Nagaty, K.: An enhanced DNA-based steganography technique with a higher hiding capacity. In: Bioinformatics, pp. 150–157 (2015)
4. Verma, V., Rishma, C.: An enhanced Least Significant Bit steganography method using midpoint circle approach. In: International Conference on Communications and Signal Processing (ICCSP). IEEE (2014)
5. Mousa, H., Moustafa, K., Abdel-Wahed, W., Hadhoud, M.M., Mousa, H.: Data hiding based on contrast mapping using DNA medium. Int. Arab J. Inf. Technol. **8**(2), 147–154 (2011)
6. Baby, D., Thomas, J., Augustine, G., George, E., Michael, N.R.: A novel DWT based image securing method using steganography. Procedia Comput. Sci. **46**, 612–618 (2015)
7. Malathi, P., Gireeshkumar, T.: Relating the embedding efficiency of LSB steganography techniques in spatial and transform domains. Procedia Comput. Sci. **93**, 878–885 (2016)
8. Wang, Z., Xiaohang, Z., Hong, W., Guangzhao, C.: Information hiding based on DNA steganography. In: 2013 4th IEEE International Conference on Software Engineering and Service Science (ICSESS). IEEE (2013)
9. Menaka, K.: Message encryption using DNA sequences. In: 2014 World Congress on Computing and Communication Technologies (WCCCT). IEEE (2014)
10. Peterson, I.: Hiding in DNA. In: Proceedings of Muse, p. 22 (2001)
11. Das, P., Nirmalya K.: A DNA based image steganography using 2D chaotic map. In: International Conference on Electronics and Communication Systems (ICECS). IEEE (2014)
12. Shiu, H.J., Ng, K.L., Fang, J.F., Lee, R.C., Huang, C.H.: Data hiding methods based upon DNA sequences. Information Sciences **180**(11), 2196–2208 (2010)
13. Fridrich, J., Soukal, D.: Matrix embedding for large payloads. IEEE Trans. Inf. Forensics Secur. **1**(3), 390–395 (2006)
14. Chan, C.K., Cheng, L.M.: Hiding data in images by simple LSB substitution. Pattern Recogn. **37**(3), 469–474 (2004)
15. Cao, Z., Zhaoxia, Y., Honghe, H., Xiangping, G., Liangmin, W.: High capacity data hiding scheme based on (7, 4) Hamming code. SpringerPlus **5**(1), 175 (2016)
16. Banoci, V., Gabriel, B., Dušan L.: A novel method of image steganography in DWT domain. In: 21st International Conference on Radioelektronika (RADIOELEKTRONIKA). IEEE (2011)
17. Meera, M., Malathi, P.: An improved embedding scheme in compressed domain image steganography. Int. J. Appl. Eng. Res. **10**(55), 1933–1937 (2015)

Dependency of Various Color and Intensity Planes on CNN Based Image Classification

Rajan Sachin(✉), V. Sowmya, D. Govind, and K.P. Soman

Centre for Computational Engineering and Networking (CEN), Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Amrita University, Coimbatore, India
insanesac2@gmail.com

Abstract. Scene classification systems have become an integral part of computer vision. Recent developments have seen the use of deep scene networks based on convolutional neural networks (CNN), trained using millions of images to classify scenes into various categories. This paper proposes the use of one such pre-trained network to classify specific scene categories. The pre-trained network is combined with the simple classifiers namely, random forest and extra tree classifiers to classify scenes into 8 different scene categories. Also, the effect of different color spaces such as RGB, YCbCr, CIEL*a*b* and HSV on the performance of the proposed CNN based scene classification system is analyzed based on the classification accuracy. In addition to this, various intensity planes extracted from the said color spaces coupled with color-to-gray image conversion techniques such as weighted average, and singular value decomposition (SVD) are also taken into consideration and their effects on the performance of the proposed CNN based scene classification system are also analyzed based on the classification accuracy. The experiments are conducted on the standard Oliva Torralba (OT) scene data set which comprises of 8 classes. The analysis of classification accuracy obtained for the experiments conducted on OT scene data shows that the different color spaces and the intensity planes extracted from various color spaces and color-to-gray image conversion techniques do affect the performance of proposed CNN based scene classification system.

1 Introduction

The task of classification occurs in a wide range of human activity. The term could broadly cover any context in which decisions have to be made on the basis of a set of available information [1,2].

Consider a given training examples of the form $\{(X_1, y_1),...,(X_n, y_n)\}$ for a function $y = f(X)$. Each X_i values can be considered as vectors of discrete values of the form $(X_{(i,1)}, X_{(i,2)},...(X_{(i,m)}))$ and are called as features of X_i. The y_i values also called as labels, and are selected from a discrete set of classes in the form $(1 \ldots K)$, where K is the number of classes. A classifier is a hypothesis of the unknown function. Given a testing examples X_t, the classifier tries to predict a corresponding y value [3].

S.M. Thampi et al. (eds.), *Advances in Signal Processing and Intelligent Recognition Systems*, Advances in Intelligent Systems and Computing 678,
DOI 10.1007/978-3-319-67934-1_15

When such tasks of classification are performed on a dataset of scenes, it is termed as scene classification, where scene can be described as something on which people can move which is greatly different from an ‘object’. The idea of scene classification has drawn a lot of attention both in academia and industry. The main objective is to automatically classify an image and assign a suitable label to the said image. Although great effort has been made, it is far from perfection yet, due to many factors such as variations in spatial position, illumination, and scale [4,5].

The increase in popularity of neural networks has aided classification and other artificial intelligence (AI) based tasks greatly. With a neural network capable of extracting features on its own to learn patterns, classification problems have attained accuracies near perfection. The use of a Convolutional Neural Networks (CNN) for classifying images has also seen great success. In CNN, a convolutional layer is normally followed by a pooling layer. Pooling is a simple method of feature extraction where the average or maximum value of a patch of neighboring features are taken and then passed on to the next layer. To create a deep convolutional neural network, multiple layers alternating between convolutional and pooling layers are stacked on each other [6,7]. This paper makes use of a pretrained CNN network and a few other classifiers like random forest classifier and extra tree classifier. A Random Forest Classifier (RF) is an ensemble [3] model, in simpler terms, a model that uses the results from many different classification models to calculate a response for one single task. A RF model creates different decision trees, typically hundreds of decision trees. Each decision tree tries to create a model for the same input train data and produces a response value. When a test data is introduced to a random forest classifier, each of the individual decision trees predicts a class accordingly to the model it had created. Finally, the class with the most number of prediction is chosen [8]. Extra Tree Classifier or Extremely Randomized Tree Classifier (ET) are slightly evolved form of random forest classifier and just add another layer of randomness to the random forest. During training a tree, instead of choosing the best split based on some optimal threshold, a randomly obtained threshold value is selected for each feature. This causes the search space to diminish thereby resulting in faster training [9].

Color space, also known as the color model (or color system), is an abstract mathematical model which simply describes the range of colors as tuples of numbers, typically as 3 or 4 values or color components (e.g. RGB). Basically speaking, color space is an elaboration of the coordinate system and sub-space. The RGB [10] data set was converted into five other color spaces, as mentioned before, that is, HSV [11,12], CIEL*a*b*(L*a*b*) [11], and YCbCr [13]. Additionally, the experiment was conducted on the L* plane of CIEL*a*b*, V plane of HSV and Y plane of YCbCr, which contains the intensity values for their own respective color spaces. Two other intensity planes were also obtained, this time by decolorizing RGB images using two different methods namely; RGB2Gray image decolorization and SVD image decolorization.

Grey scale or more commonly known as black and white images are different from RGB images in the aspect that, instead of 3 color planes, grayscale images are made of a single color plane. Each pixel in the grayscale images range from 0 to 255, where 0 represents black and 255 represents white and any number in between are shades of either black or white [14]. The three different colors have three different wavelengths and have their own contribution to the formation of an image. Of all the three color red has the most wavelength. Green, on the other hand, is the most sensitive to human eyes. And so green is given the most weight age.

$$Greyscale = (0.3 * R + 0.59 * G + 0.11 * B) \tag{1}$$

In the second technique for color-to-grayscale image conversion [15], chrominance information from the color images are incorporated into the luminance information to obtain a gray scale image. The chrominance information is reconstructed using the eigenvectors and eigenvalues obtained through singular value decomposition (SVD) in the CIEL*a*b* color space instead of RGB space, this is because color images in CIEL*a*b* enables processing the luminance and chrominance components independently. In this paper, this version of the dataset is also referred to as SVD data set for ease of use.

So a total of 4 color spaces and 5 intensity planes for made use of to study how the accuracy of the classification task is affected. The experiment was conducted on OT data set that comprises of 2688 RGB images of outdoor places belonging to 8 different classes where each image of 256 × 256 pixels in size [16]. Figure 1 shows nine versions of the image ‘coast5.jpg’.

Hence, this paper analyzes the effect of different color space models on scene classification. Along with this, the paper also analyzes the effect two different decolorization techniques, rgb2gray and SVD image decolorization technique on

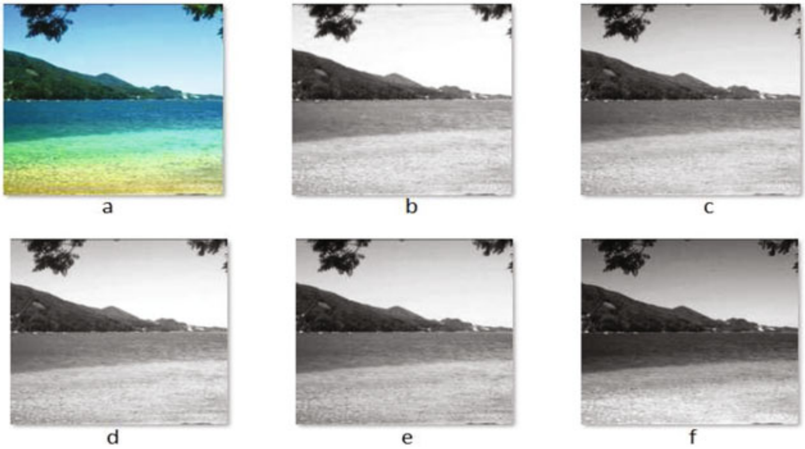


Fig. 1. (a) RGB, (b) V plane, (c) L plane, (d) Y plane, (e) RGB2Gray (f) SVD decolorized image

the same classification task in comparison to the effect of different intensity planes. For the purpose of classification, this paper proposes a pretrained CNN employed together with random forest (RF) and extra tree (ET) classifiers. The scene classification task was then performed on OT scene data which comprises of 8 classes.

This paper is organized as follows: the methodology is discussed in Sect. 2 while the experimental results and analysis are covered in Sect. 3. Section 4 gives the summary and the conclusion for this paper, finally followed by the related references.

2 Methodology

The first step involved feeding the RGB data set to the pretrained CNN model. The comparatively small data set made it hard to design a new neural network that could give satisfactory accuracy and hence a predefined model was chosen. But before the dataset was fed into the CNN, the data set had to be split into training and testing set. The RGB data set was split into training and testing set, comprising of 1888 and 800 images respectively. Table 1 shows the split of data into training and testing set. The flow chart is provided in Fig. 2.

Table 1. Training and testing split of OT data used for the present work

Class	Open Country (class 0)	Coast (class 1)	Forest (class 2)	Highway (class 3)	Inside City (class 4)	Street (class 5)	Mountain (class 6)	Tall buildings (class 7)
Training set	310	260	228	160	208	274	192	256
Testing set	100	100	100	00	100	100	100	100

The images were then renamed according to their classes, for example, coast1, coast2 etc., to facilitate easy labeling. The classes were designated with labels from 0 to 7.

Due to the small size of data set, training a CNN for the purpose of classification is really hard. And hence, the experiment was done using a pretrained network. For the purpose of training Places-CNN, 2,448,873 images from 205 categories of Places-205 dataset were selected randomly as the train set, with minimum 5,000 and maximum 15,000 images per category. The validation set was made of 100 images per category while the test set contained 200 images per category to give a total of 41,000 images. Places-CNN model was trained using the Caffe package on a NVIDIA Tesla K40 GPU and took about 6 days to finish 300,000 iterations of training [17]. The network contains eight layers with weights; five convolutional and three fully-connected layers. The output of the last fully-connected layer produces a distribution over the 205 class labels.

The input layer for the network is an image of size 224 × 244 × 3, which is then fed to a convolutional layer with 96 kernels of size 11 × 11 and a stride of 4 pixels. To the output of every convolutional and fully connected layer, a ReLU nonlinearity is applied to. Each neuron computes the weighted sum of its inputs and applies an offset which then runs the result through a nonlinear function. ReLU is a nonlinear function that results in the faster training of neural networks.

The first, second and fifth convolutional layers are followed by max pool layers. Response-normalization layers follow the first and second convolutional layers. The output of the first convolutional layer (after max pooling and normalization) acts as the input to the second convolutional layer, which then takes the input and filters it with 256 kernels of size 5 × 5 × 48 and also adds a padding of 2. The third and fourth convolutional layers are void of any pooling or normalization layers so is the connection between fourth and fifth convolutional layers. The third convolutional and fourth convolutional layers have 384 kernels of size 3 × 3 × 256 and 384 kernels of size 3 × 3 × 192. The fifth and last convolutional layer accepts the output from the fourth convolution layer as input and has 256 kernels of size 3 × 3 × 192.

All the fully-connected layers have 4096 neurons each, but only the first and second fully-connected layers have a drop out layer, with a dropout factor of 0.5, following them. The final fully-connected layer is then fed to 205 way softmax to obtain prediction on which class the image belongs to [18]. For this paper, such a prediction of the class is not necessary. Instead, for the purpose of the experiment, a series of 205 values, where each value denotes the probability of a particular image falling into one particular class, is required. And so, a slight change to the network is made, that is, the exclusion of the softmax function thus making the CNN provide a vector of size 1 × 205 for each image, instead of a particular class, thus proposing a slightly altered network.

The pretrained model predicted, out of the 205 classes, to which class the new observation belongs to. For the purpose of this experiment, rather than predicting a particular class, the model was made to give an output of 205 values. This step was necessary because, while the dataset we used had specific classes, for example, coast, the 205 classes had classes like coast, pond, aquarium and so on. To simplify it, an image from the class coast was categorized to coast, pond, aquarium which all had a common factor, water.

The model was then fed with the training and testing data set separately. The training set comprised of 1888 images, that means for each image, the model gave out 205 probabilistic values for each image. These 205 values when added resulted into 1, in other words, these values were probabilistic values. In a classification task, generally, the likelihood of a particular image falling into all the classes is calculated by the CNN, and then the class with the highest probability or likelihood value is then considered as the class in which the image belongs. Here, since there are 205 classes, the CNN will calculate 205 probability values for every single image. Instead of choosing the highest probability value, we made use of the entire 205 values to form a feature vector which was then

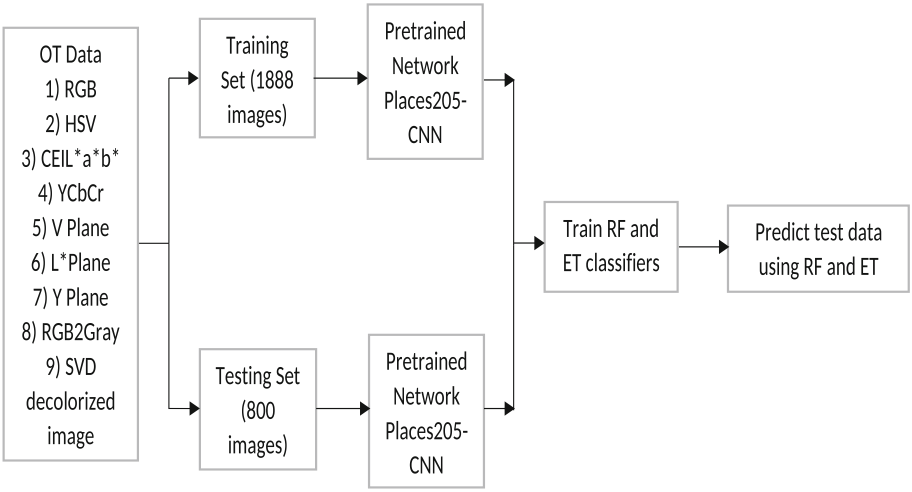

OT Data
1) RGB
2) HSV
3) CEIL*a*b*
4) YCbCr
5) V Plane
6) L*Plane
7) Y Plane
8) RGB2Gray
9) SVD decolorized image
Training Set (1888 images)
Testing Set (800 images)
Pretrained Network Places205-CNN
Pretrained Network Places205-CNN
Train RF and ET classifiers
Predict test data using RF and ET

Fig. 2. Flow chart for the proposed scene classification system

used for further classification. Thus a matrix of shape 1888 by 205 was obtained, each vector of size 205 represented one image. The same was done with the testing set to obtain a matrix of shape 800 by 205.

Table 2. Accuracies (%) obtained for RGB dataset using the training matrix

Classifiers	Linear SVM	Decision tree	Gradient boost	Random Forest (RF)	Extra tree	MLP	CNN
Overall accuracy	83.62	82.91	89.06	90.02	90.02	81.85	87.68

The next step involved training classifiers using the training set to obtain a prediction on the testing set. A number of classifiers namely linear SVM, decision tree, RF, ET, multi layer perceptron (MLP) and a new CNN network were used to train and test on the two matrices obtained.

These classifiers were made use to map the 205 classes into the original 8 classes. A true label containing values from 0 to 7 was generated where images of the same class were given one of the values, for example, 0 for opencountry, 1 for coast and so on. This process was done for both the training and testing set to get a training label and a testing label. With the help of the training matrix and the training label, 6 new classifiers were trained to map 205 values into 8 classes. Once the training was done, these 6 classifiers were used to predict the classes for the testing matrix and the accuracies were then compared. Out of the various classifiers considered, only two classifiers, random forest and extra tree classifier, that produced the highest overall accuracies were then chosen for further analysis.

The RGB data set was also duplicated 5 times, with each duplicated dataset undergoing a transformation in color space. One of the datasets was converted into the HSV color space, another to CEIL*a*b* and the third one into YCbCr. The remaining two copies of the RGB dataset were converted into grayscale datasets using RGB2Gray image decolorization and SVD image decolorization techniques respectively. The same experiment explained above was then conducted on all the five datasets separately and the results were tabulated.

3 Result and Analysis

Table 2 shows the various classifiers that were initially chosen and their respective accuracies. As evident from Table 2, RF and ET outperforms all the other classifiers, by obtaining an overall accuracy of 90.02%. And hence, these two were chosen for further experiments. As mentioned earlier, the classification experiment was conducted on 5 intensity planes including the two decolorized version and 4 color spaces. For each experiment, a random forest classifier, as well as an extras classifier, was trained using the same data set and these classifiers were made to predict the same training set. The experiment was repeated several times, in order to obtain maximum accuracy, by fine tuning these said classifiers. In the case of the random forest classifier, the parameter tuned was the number of trees, also called 'estimators'. While for extra tree classifiers, the parameters to be tuned were, the number of estimators as well as the number of features to consider when looking for the best split. Here, accuracy is defined as [19]:

Accuracy = (Number of correctly predicted test images/Total number of test images) * 100

Table 3 summarizes the overall accuracies for all the different color and intensity planes for RF and ET classifiers. As evident from the table, extra tree classifier tends to have better accuracy in most cases, except for HSV and RGB color space. With an accuracy of 78.12%, the random forest slightly outperforms the extra tree classifier, by a margin of 2.0% for HSV images. As for RGB, the ET classifier produces an accuracy of 90.0% which falls shorter by 0.125% as compared to the 90.12% produced by RF. In all the other cases, the ET slightly outperforms the random forest classifier. When we compare the results of the color spaces with each other, rather than the classifiers, HSV-based image classification, with an accuracy of 78.12% and 76.12% for RF and ET respectively, lags behind the other color spaces by a huge margin. When compared to CIEL*a*b*, the which ranks immediately above HSV in terms of accuracy, a difference of 6.25% and 9.62% for the classifiers is evident. On the other hand, YCbCr and CIEL*a*b* shares almost similar accuracies. The individual planes, V, L* and Y, which are intensity values, just like RGB2Gray decolorized images, all share similar accuracies. In fact, their accuracies are similar to that of grayscale images. The highest accuracy was obtained for the grayscale version which incorporated chrominance information, with an accuracy of 91.24% for RF and 91.62% for ET.

Table 3. Overall accuracy (%) for the different color and intensity planes of the OT dataset

Color space	RGB	HSV	CIELAB	YCbCr	V Plane	L *Plane	Y Plane	RGB2Gray	SVD decolorized
RF	90.02	77.87	84.37	84.62	90.50	90.87	90.75	90.62	91.37
ET	90.02	77.62	85.62	86.50	91.00	91.37	91.00	91.25	91.62

When compared to the immediately below accuracies of the RGB2Gray decolorized images, even though the difference is not vast, it is certainly a significant improvement in the case of a classification problem.

Table 4 summarizes the class wise accuracy for each color space for RF and ET classifiers. Comparing the data, the most number of misclassification happens in class 6 (mountains) followed by class 0 (opencountry). Both these classes comprise of images with a large region of sky. And so the classifiers gets confused and hence the misclassified. Some images in class 0 are hilly in natures which again contributes to the misclassification. On the other hand, the least number of misclassification occurs in class 7 (tallbuildings), with an average accuracy of 96.11% for both RF and ET.

Table 4. RF and ET class wise accuracy (%) for different color and intensity planes of the OT dataset

Color space	Classifier	class 0	class 1	class 2	class 3	class 4	class 5	class 6	class 7
RGB	RF	84	93	94	90	96	88	80	97
	ET	82	91	92	93	96	88	83	97
HSV	RF	70	68	84	73	79	93	76	90
	ET	69	71	85	68	75	86	80	87
CIEL*a*b*	RF	79	80	95	82	82	82	80	95
	ET	82	82	96	84	85	80	79	97
YCbCr	RF	70	58	91	80	87	86	83	95
	ET	75	87	93	82	87	86	86	96
V Plane	RF	84	95	93	90	97	83	82	98
	ET	84	95	95	91	96	87	82	98
L* Plane	RF	83	94	97	90	99	84	82	98
	ET	87	94	94	92	97	88	82	97
Y Plane	RF	86	92	97	92	97	85	80	97
	ET	86	95	96	93	94	87	80	97
RGB2Gray	RF	85	93	96	92	97	85	80	97
	ET	84	98	95	93	96	87	79	98
SVD decolorized	RF	84	94	97	92	99	84	83	98
	ET	89	95	97	91	97	85	81	98

Class 5 (Street) from every dataset scored comparatively low as compared to the other classes. With exception of a single case (93 correct predictions for HSV dataset), the classifiers could only correctly predict 80 to 88 images. This might be due to the similarities with class 3 (highway), class 4 (insidecity) and class 7 (tall buildings).

In the case of the RGB data set, both RF and ET had correctly classified 722 images out of 800. Even though both classifiers have the same accuracy, the class wise accuracy varies widely. The class with most misclassification for RF is class 6 while for ET is class 0. When comparing the class wise accuracy of YCbCr space, the most misclassification happens in class 6. The ET classifier was able to get a testing accuracy of 86% the said class, 5% more when compared to the ET of SVD decolorized image data set. Both ET and RF were able to predict 88% correct classification for class 5 (street) for the RGB version, which is better than all the other version except for the L* plane table.

Comparing the class wise accuracy for HSV set and chrominance data set (the lowest and highest accuracy color planes), all the classes have better accuracy for the images that were SVD decolorized, except for class 5. On comparison, both RF and ET classifiers were able to provide a better prediction for the HSV version. In fact, the RF registered 93% accuracy, which is better than all the other cases.

The ET classifier had an accuracy of 89% with the SVD data set for class0. Due to its nature of images, the open country images from class0 have chances of being considered as a coast or as a mountain, which has led to comparatively high misclassifications.

The three individual plane data set are kind of similar, as the pixel values are just intensity values like that of gray scale. All the four cases share an almost similar over all accuracy for both RF and ET, but when the class wise accuracy is considered, the same is not true. The classes with the least misclassification might be same for all these four versions, but the number of misclassification slightly differ.

After a detailed comparison of the ET classifier, on all the tables, class 0, class 2, class 4 and class 7 have the least misclassification for the gray scale with chrominance information data set, weighted grayscale version for class 1 (coast), RGB data set for class 3 (highway) and class 5, and finally YCbCr for class 6. When a similar analysis is done for the RF classifier, Y plane data set gave better accuracy for class 0 and V plane data set for class 1. As for class 2 (forest), the gray scale with chrominance, L* plane and Y plane all shared an accuracy of 97%. The RGB2Gray and SVD data set was the best suited for class 3. For class 4 and class 5, the chrominance data set and HSV data set showed better results. As for class 6, once again the chrominance data set outperformed the other cases. And finally for class 7, V plane version, L* version and also the chrominance version had the least misclassification of 2 images.

4 Conclusion

The present work proposes the utilization of a pretrained network, Places205-CNN to classify a scene data set comprised of 8 classes. The RGB data set was converted to 3 different color spaces and 5 intensity planes, to form a total of 9 versions of the dataset. The dataset was split into training and testing set and fed into the network. For every image, a feature vector of dimension 1 × 205 was produced as output by the network. A RF and an ET classifier were then trained using the feature matrix of the training set and predictions were made on the feature matrix of the testing set. The overall accuracy and class wise accuracy were compared and analyzed.

On analyzing the various overall and class wise accuracies, it was understood that the color spaces and intensity planes did affect the overall and class wise accuracies of the classification task. All the intensity planes, though similar, produced variations in the overall and class wise accuracy. We also concurred that different classes were predicted better in a particular color space or intensity plane.

This leads to the possibility that, if a classifier were trained on a dataset with classes belonging to various color spaces, the overall accuracy and class wise accuracy of deep scene classification system might improve class wise as well as the overall accuracy. Thus, the effects of color spaces on classification are not negligible.

References

1. Megha, P., Swarna, M., Dixon, D., Sowmya, V., Soman, K.: Impact of least square denoising on kernel based hyperspectral image classification. Int. J. Control Theor. Appl. **9**(10), 4623–4630 (2016)
2. Michie, D., Spiegelhalter, D.J., Taylor, C.C.: Machine learning, neural and statistical classification (1994)
3. Dietterich, T.G.: Ensemble methods in machine learning. In: International Workshop on Multiple Classifier Systems, pp. 1–15. Springer (2000)
4. Zou, J., Li, W., Chen, C., Du, Q.: Scene classification using local and global features with collaborative representation fusion. Inf. Sci. **348**, 209–226 (2016)
5. Dutt, B.S.R., Agrawal, P., Nayak, S.: Scene classification in images (2009)
6. Raschka, S.: Python Machine Learning. Packt Publishing Ltd., Birmingham (2015)
7. Athira, S., Rohit, M., Prabaharan, P., Soman, K.: Automatic modulation classification using convolutional neural network. Int. J. Comput. Technol. Appl. **9**(16), 7733–7742 (2016)
8. Horning, N., et al.: Random forests: an algorithm for image classification and generation of continuous fields data sets, New York (2010)
9. Maier, O., Wilms, M., von der Gablentz, J., Krämer, U.M., Münte, T.F., Handels, H.: Extra tree forests for sub-acute ischemic stroke lesion segmentation in mr sequences. J. Neurosci. Methods **240**, 89–100 (2015)
10. Basha, M.S., Ramakrishnan, M.: Color image contrast enhancement using daubechies d4 wavelet and luminance analysis. Int. J. Comput. Appl. **86**(6), 6–10 (2014)

11. Plataniotis, K., Venetsanopoulos, A.N.: Color Image Processing and Applications. Springer Science & Business Media, Heidelberg (2013)
12. Wen, C.-Y., Chou, C.-M.: Color image models and its applications to document examination. Forensic Sci. J. **3**(1), 23–32 (2004)
13. Kaur, A., Kranthi, B.: Comparison between ycbcr color space and cielab color space for skin color segmentation. IJAIS **3**(4), 30–33 (2012)
14. Wu, T., Toet, A.: Color-to-grayscale conversion through weighted multiresolution channel fusion. J. Electron. Imaging **23**(4), 1–6 (2014)
15. Sowmya, V., Govind, D., Soman, K.: Significance of incorporating chrominance information for effective color-to-grayscale image conversion. Sign. Image Video Process. **11**(1), 129–136 (2017)
16. Oliva, A., Torralba, A.: Modeling the shape of the scene: a holistic representation of the spatial envelope. Int. J. Comput. Vis. **42**(3), 145–175 (2001)
17. Zhou, B., Lapedriza, A., Xiao, J., Torralba, A., Oliva, A.: Learning deep features for scene recognition using places database. In: Advances in Neural Information Processing Systems, pp. 487–495 (2014)
18. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems, pp. 1097–1105 (2012)
19. Jiang, S., Liu, D.: On chance-adjusted measures for accuracy assessment in remote sensing image classification. In: ASPRS Annual Conference (2011)

Epigraphic Document Image Enhancement Using Retinex Method

H.T. Chandrakala[1(✉)] and G. Thippeswamy[2]

[1] Visweswaraya Technological University, Bengaluru, Karnataka, India
chandrakl80@gmail.com
[2] BMS Institute of Technology, Bengaluru, Karnataka, India
swamy.gangappa@gmail.com

Abstract. Epigraphic Documents are the ancient handwritten text documents inscribed on stone, metals, wood and shell. They are the most authentic, solitary and unique documented evidences available for the study of ancient history. In the recent years, Archeological Departments worldwide have taken up the massive initiative of converting their repository of ancient Epigraphic Documents into digital libraries for the perennial purpose of their preservation and easy dissemination. The visual quality of the digitized Epigraphic Document images is poor as they are captured from sources that would have suffered from various kinds of degradations like aging, depositions and risky handling. Enhancement of these images is an essential prerequisite to make them suitable for automatic character recognition and machine translation. A new approach for enhancement of Epigraphic Document images using Retinex method is presented in this paper. This method enhances the visual clarity of the degraded images by highlighting the foreground text and suppressing the background noise. The method has been tested on digitized estampages of ancient stone inscriptions of 11th century written in old Kannada language. The results achieved are efficient in terms of root mean square contrast and standard deviation.

Keywords: Epigraphic documents · Single scale retinex · Multi scale retinex · Gaussian surround

1 Introduction

Epigraphic documents which are available numerously worldwide, are the most primary and authentic sources of the social, cultural, economic, administrative and dynastic history of mankind. These documents can be dated back to ancient time periods when writing material like pen and paper were not yet available. Unfortunately, these valuable sources of written history are at the verge of extinction due to various forms of degradations like aging, harsh weather conditions, natural disasters, dust, deposition and risky handling. As a step forward towards preservation and propagation of these valuable possessions to the future generations, Archeological departments throughout the world create their true copies in the form of estampages and store them in their corpus. But estampages will also spoil over the years due to breakages and wear and tear.

S.M. Thampi et al. (eds.), *Advances in Signal Processing and Intelligent Recognition Systems*, Advances in Intelligent Systems and Computing 678,
DOI 10.1007/978-3-319-67934-1_16

Digitization is a more reliable technological solution for increasing the shelf life of estampages. Estampages are camera captured or scanned and stored as digital images as part of the digitization process. Owing to digitization Epigraphic documents can become easily searchable and accessible to the public which was not possible otherwise. Also, it opens the scope to automate the mundane tasks of era identification and transliteration of these documents written in ancient script into modern language. But the main constraint for such automatic machine recognition and translation is the poor visual quality of the digitized epigraphic documents. Thus, it is inevitable to enhance these images to improve their visual clarity. Image Enhancement is a step which highlights the text contours which are obscured and reduces the background noise, thus making them more suitable for further image processing steps like segmentation, feature extraction and Optical Character Recognition (OCR).

2 Related Work

Image Enhancement approaches in use presently can be broadly classified into spatial domain approach and frequency domain approach. Intensity transformation, filtering [5, 9, 10, 12, 18] Independent Component Analysis(ICA) and histogram equalization [20] are the methods commonly used in spatial domain. In frequency domain, discrete wavelet transform [14, 17, 19], curvelet transform [2], shearlet transform [16] are commonly used for contrast enhancement of images. But these techniques are not adequate for epigraphic document images which suffer from severe degradations and lack of sharp separation between foreground text and the background. Such degradations can be effectively handled by Retinex method which can enhance the image contrast making use of local pixel information. Retinex [4, 6, 7, 8] can perform contrast stretching and dynamic range compression simultaneously thus providing superior visual quality for the image. This technique has been applied for enhancement of natural scene images [1], microscopic images [3], satellite images [11] and medical images [15]. Retinex method for enhancement of Document images has been attempted for the very first time in our work.

This paper outlines two variants of the Retinex technique namely: Single Scale Retinex and Multi Scale Retinex and discusses their applicability for enhancement of Epigraphic Document images. The rest of the paper is organized as follows: Sect. 2 introduces the Retinex concept, Sect. 3 gives a detailed explanation of the proposed enhancement scheme, Sect. 4 discusses the experimental results and discussion and Sect. 5 concludes the paper.

3 Retinex Theory

Retinex theory which models how the human visual system perceives a scene was proposed for the first time by Land and McCann [8, 13] in their US patent. They established that the lightness perceived by the human visual system is rather relative as opposed to absolute. Retinex computes the lightness values of an image that will be invariant under changes of viewing context, just like human vision is roughly invariant under similar change. According to basic image formation model, every image F (x, y) is made up of two essential ingredients: Illumination I (x, y) and Reflectance R (x, y) which is mathematically represented as:

$$\mathrm{F}\ (x, y) = \mathrm{I}\ (x, y)\ *\ \mathrm{R}\ (x, y) \tag{1}$$

The Retinex algorithm compensates for non-uniform lighting in a given image by separating its illumination component from the reflectance. It decreases the influence of the reflectance component, thus enhancing the original image to its true likeness.

4 Proposed Method

Based on the image formation model, the basic form of Single Scale Retinex (SSR) [15] is given by:

$$\mathrm{R_i}(x, y) = \log \mathrm{I_i}(x, y) - \log(\mathrm{I_i}(x, y)) * \mathrm{F}\ (x, y) \tag{2}$$

Where $\mathrm{R_i}(x, y)$ is the Retinex output, the subscript i represents the three different color channels R, G, B; $\mathrm{I_i}(x, y)$ is the i-th image of the spectrum zone; * denotes the convolution operation; F (x, y) refers to normalized surround Gauss function given by:

$$\mathrm{F}\ (x, y) = Ke^{-r^2/c^2} \tag{3}$$

$$r = \sqrt{x^2 + y^2} \tag{4}$$

c is the Gaussian surround space constant; K is a normalized parameter. So that $\iint \mathrm{F}(x, y)dxdy = 1$

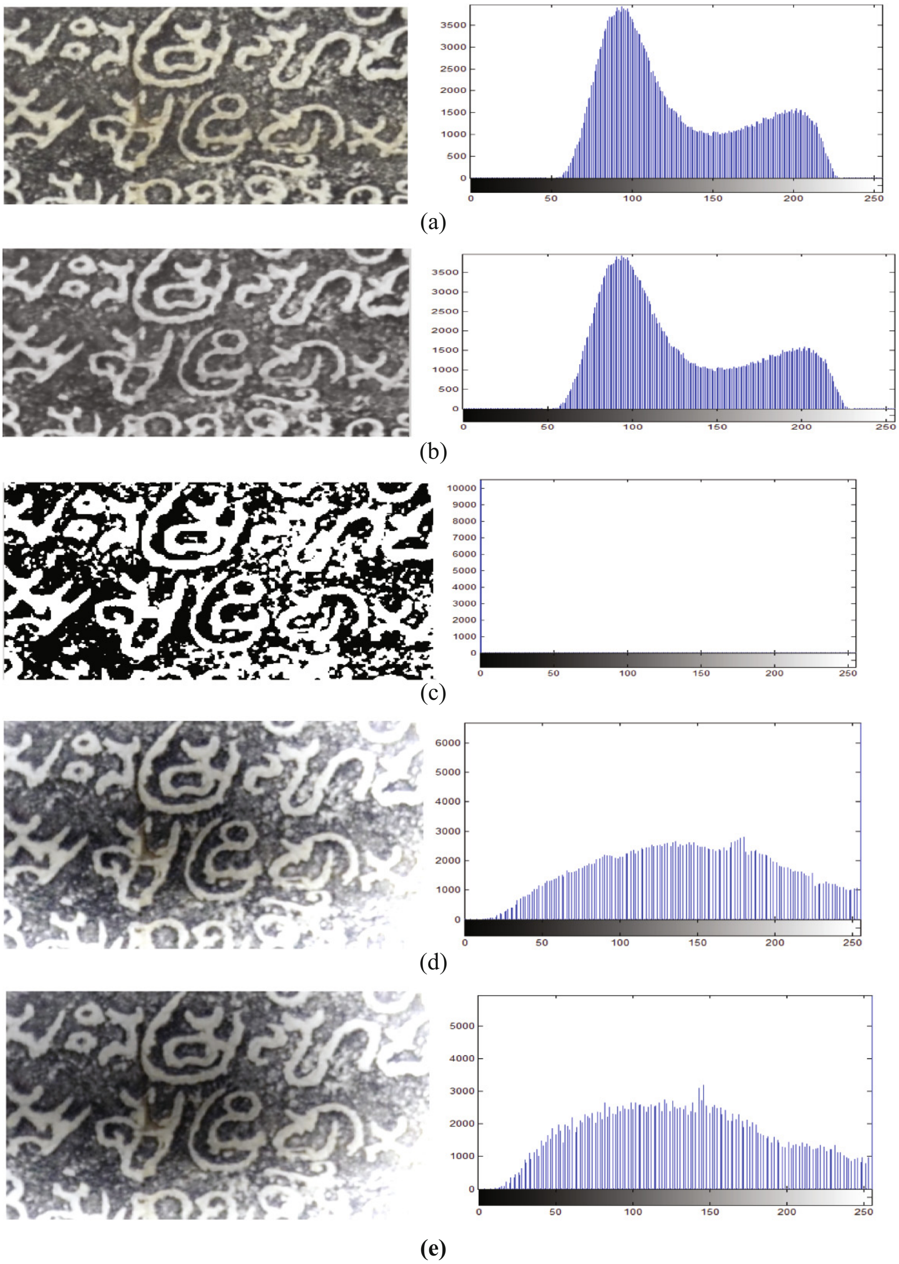
The Multi-Scale Retinex (MSR) is an expanded system of SSR. This method is realized by the weighted sum of different SSR outputs which are processed by different scale Gaussian functions. The MSR algorithm is defined by:

$$R_{MSR_i} = \sum\nolimits_{n=1}^{N} \omega_n R_{n_j} \tag{5}$$

Where N is the number of scales; R_{n_j} is the result of the i-th spectrum zone of the n-th scale; R_{MSR_i} is the result of MSR algorithm for the i-th spectrum zone; ω_n is the weighting coefficient of the n-th scale. To preserve the most information $\mathrm{R_i}(x, y)$ must be auto-stretched before being displayed.

5 Experimental Results and Analysis

A standard dataset of Kannada estampage data is not yet available, so a dataset of 200 camera captured images of ancient Kannada inscription Estampages that belong to the Kalyani Chalukyan era of 11th century was created to test the proposed Enhancement scheme. Figure 1 shows a comparison of the results of enhancement and their



(a)

(b)

(c)

(d)

(e)

Fig. 1. (a) Original image (b) Median Filter (c) ICA (d) SSR (e) MSR corresponding histogram plots obtained using Median Filter, Independent Component Analysis (ICA), Single Scale Retinex (SSR) and Multi Scale Retinex (MSR) techniques. The X axis of the histogram plot represents the pixel value and Y axis the number of pixels. The quality of the enhancement results is evaluated by measuring their Standard Deviation and Root Mean Square (RMS) Contrast.

Standard Deviation

$$\text{STD} = \sqrt{\frac{1}{N-1}\sum\nolimits_{i=1}^{N}\left(X_i - \overline{X}\right)^2} \tag{6}$$

where X_i is a one-dimensional array of N pixel intensities of the given image and $\overline{X}$ is the corresponding mean given by:

$$\overline{X} = \frac{1}{N}\sum\nolimits_{i=1}^{N} X_i \tag{7}$$

RMS Contrast

$$\text{RMS contrast} = \sqrt{\frac{1}{MN}\sum\nolimits_{i=0}^{N-1}\sum\nolimits_{j=0}^{M-1}\left(I_{ij} - \overline{I}\right)^2} \tag{8}$$

where I_{ij} is an image of size M × N whose pixel intensities are normalized in the range [0,1]. $\overline{I}$ is the mean intensity of all pixel values in the image I_{ij}.

Table 1 shows that Retinex techniques produce better results in terms of RMS contrast and Standard Deviation than the traditional median filtering approach.

Table 1. RMS Contrast and Standard Deviation for different Algorithms

Image	RMS Contrast	Standard Deviation
Original	2.80	44.70
Median Filter	2.80	44.71
ICA	4.02	63.90
SSR	4.60	75.79
MSR	4.58	76.53

6 Conclusion

The paper presented Single-Scale Retinex and Multi-Scale Retinex approach for the enhancement of Epigraphic Document Images. The weights associated with each SSR output image are computed according to the degradation characteristics and pixel value of the input image to get high contrast image with even tonal rendition for the entire image. It is evident from the experimental results that Retinex approach can efficiently highlight the text contours and suppress the background noise as it considers pixel level information. RMS contrast and standard deviation of the Retinex enhanced images are found to be better than the median filtering and ICA approaches which have been used earlier for Epigraphic document image enhancement. However, Retinex enhanced images suffer from slight graying effect in some parts due to over enhancement. Overcoming this problem would be considered as a future work.

Acknowledgement. We thank the organization, Archeological Survey of India(ASI), Mysore for providing access to their corpus of ancient Kannada inscription Estampages belonging to the Kalyani Chalukyan era of 11th century to conduct our research.

References

1. Petro, A.B., Sbert, C., Morel, J.-M.: Multiscale Retinex, Image Processing Online (IPOL) (2014). ISSN 2105-1232
2. Gangamma, B., Srikanta Murthy K.: A combined approach for degraded historical documents denoising using curvelet and mathematical morphology. In: Proceedings of International Conference on Computational Intelligence and Computing Research. IEEE (2010). ISBN 978-1-4244-5967-4/10
3. Biswas, B., Roy, P., Choudhuri, R., Sen, B.K.: Microscopic image contrast and brightness enhancement using multi-scale Retinex and cuckoo search algorithm. Procedia Comput. Sci. **10**, 348–354 (2015). Elsevier
4. Funt, B., McCann, J.: Retinex in Matlab. J. Electron. Imaging **V13**(I), 48–57 (1999)
5. Yuan, C., Li, Y.: Switching median and morphological filter for impulse noise removal from digital images. J. Optik **126**, 1598–1601 (2015). Elsevier
6. Land, E.H.: The Retinex Theory of color vision. J. Sci. Am. **237**(6), P108–P128 (1997)
7. Land, E.H.: Recent advances in Retinex theory. Vision. Res. **26**(1), 7–21 (1986)
8. Land, E., McCann, J.: Lightness and Retinex theory. J. Optical Soc. America **61**(1), 1 (1971)
9. Bhuvaneswari, G. Subbiah Bharathi, V.: An efficient algorithm for recognition of ancient stone inscription characters. In: Proceedings of 7th International Conference on Advanced Computing. IEEE (2015). ISBN 978-5090-1933-5/15
10. Janani, G., Vishalini, V., Mohan Kumar, P.: Recognition and analysis of tamil inscriptions and mapping using image processing techniques. In: Proceedings of Second International Conference on Science Technology Engineering and Management. IEEE (2016). ISBN 978-1-5090-1706-5/16
11. Hines, G., Rahman, Z.U., Jobson, D., Wbodell, G.A.: Single-scale retinex using digital signal processors. In: NASA Research Report, Proceedings of Global Signal Processing Conference (2005)
12. Sreedevi, I., Pandey, R. Jayanthi, N., Bhola, G., Chaudhary, S.: Enhancement of inscription images. In: Proceedings of National Conference on Communications. IEEE (2013). ISBN 978-4673-5952-8/13
13. Frankle, J., McCann, J.: Method and apparatus of lightness imaging, US Patent #4,384,336 (1983)
14. Wang, Q., Xia, T., Li, L., Tan, C.L.: Document image enhancement using directional wavelet. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. IEEE (2003). 1063-6919/03
15. Meng, Q., Bian, D., Guo, M., Lu, F., Liu, D.: Improved Multiscale Retinex Algorithm For Medical Image Enhancement Information Engineering and Applications. Springer-Verlag, London (2012). doi:10.1007/978-1-4471-2386-6_121

16. Ranganatha, D., Holi, G.: Historical document enhancement using shearlet transform and mathematical morphological operations. In: Proceedings of International Conference on Advances in Computing, Communications and Informatics. IEEE (2015). ISBN 978-1-4799-8792-4/15
17. Pasha, S., Padma, M.C.: Handwritten kannada character recognition using wavelet transform and structural features. In: Proceedings of International Conference on Emerging Research in Electronics, CST. IEEE (2015). ISBN 978-4673-9563-2/15
18. Soumya, A., Hemantha Kumar, G.: Enhancement and segmentation of historical records. In: Computer Science & Information Technology (CS & IT) Computer Science Conference Proceedings (CSCP), vol. 15 (2015). ISSN 2231-5403
19. Yan, C.C.: Image Enhancement by adjusting the contrast of spatial frequencies. Optik J. **119**, 143–146 (2008)
20. Jin, Y., Fayad, L.M., Laine, A.F.: Contrast enhancement by multiscale adaptive histogram equalization. Proc. SPIE **4478**, 206–213 (2001)

Improved Microaneurysm Detection in Fundus Images for Diagnosis of Diabetic Retinopathy

V. Dharani(✉) and R. Lavanya

Department of Electronics and Communication Engineering,
Amrita School of Engineering, Coimbatore, Amrita Vishwa Vidyapeetham,
Amrita University, Coimbatore, India
dharani.vasu213@gmail.com, r_lavanya@cb.amrita.edu

Abstract. This paper addresses the development of a computer-aided diagnosis (CAD) system for early detection of diabetic retinopathy (DR), a sight threatening disease, using digital fundus photography (DFP). More specifically, the proposed CAD system is intended for detection of microaneurysms (MA) which are the earliest indicators of DR; CAD systems for MA detection involve two stages: coarse segmentation for candidate MA detection and fine segmentation for false positive elimination. The system addresses the common challenges in candidate MA detection, which includes detection of subtle MAs and MAs close to each other and those close to blood vessels which leads to low sensitivity. The system employs four major steps. The first step involves preprocessing of the fundus images, which comprises of shade correction, denoising and intensity normalization. The second step aims at the segmentation of candidate MAs using bottom hat transform, thresholding and hit or miss transformation. The use of modified morphological contrast enhancement and multiple structuring elements (SEs) in the hit or miss transform has improved the detection rate of MAs. The proposed method has been validated using a set of 20 fundus images from the DIARETDB1 database. The Free Response Operating Characteristics (FROC) curve demonstrates that many MAs that are otherwise missed out are detected by the proposed CAD system.

Keywords: Microaneurysm detection · Computer-aided diagnosis · Diabetic retinopathy · Normalization · Shade correction · Structuring elements

1 Introduction

Diabetic retinopathy (DR) is the most common complication of diabetes mellitus (DM) characterized by abnormal or damaged blood vessels in the retinal structure of the eye. It is one of the major causes of blindness in people of 20−65 years of age [1]. Approximately 382 million people across the world have been estimated to have DM in 2013 and this can rise to 592 million by 2035 [2]. After the onset of DM, there is increased chance for developing DR over the years. DR is asymptomatic and goes unnoticed until it reaches the advanced stage, and it is necessary to do a timely diagnosis with the help of better screening options and facilities [3]. Early diagnosis of DR helps in prevention of vision loss and impairment.

S.M. Thampi et al. (eds.), *Advances in Signal Processing and Intelligent Recognition Systems*, Advances in Intelligent Systems and Computing 678,
DOI 10.1007/978-3-319-67934-1_17

Based on the development of pathological features, DR is broadly classified into non-proliferative DR (NPDR) and proliferative DR (PDR). Various clinical features present through the different stages of DR are: Microaneurysm (MA), haemorrhages (HEM), hard exudates, soft exudates, neovascularisation and macular edema. NPDR occurs first and PDR is the advanced stage where there is development of new abnormal blood vessels. The treatment options available at the stage of PDR such as laser photocoagulation, anti-VEGF injection, intraretinal injection and vitrectomy are found to be less effective and do not provide the recovery of vision loss that has already taken place [2]. MAs are small protrusions within the capillary walls that appear as minute red dots on the retinal surface of the eye, and start to develop at the NPDR stage.

MAs are the first visible sign of DR [4]. MAs are low contrast circular structures with size ranging from 10 μm to 100 μm, usually less than 125 μm [5]. They share similar characteristics with other anatomical features such as HEM and blood vessels. It is necessary to extract MAs from other MA like structures for proper diagnosis and staging of DR. The severity of disease is indicated by the number of MAs as shown in Table 1.

Table 1. Grading of DR [5]

Sl.No	Stage of DR	No/Type of lesions
1	Grade 0	MA = 0;
2	Grade 1	$1 \leq$ MA ≤ 5;
3	Grade 2	5 < MA <15;
4	Grade 3	MA ≥ 15;

It is evident from Table 1 that accurate detection of MA without overlooking them is essential for accurate staging of DR, which in turn is used for appropriate diagnosis and treatment options. Early detection of MAs can help in prevention of vision loss. People who are affected with DM must undergo regular screening to diagnose MAs at an early stage. For screening programs for a large population which involves relatively fewer expert ophthalmologists, a computer aided diagnosis (CAD) system can reduce the cost and workload involved. This works aims at the detection of MAs, using digital fundus photography (DFP) with emphasis on not missing out the difficult cases that include subtle MAs and those that are close to each other and proximal to the blood vessels.

This paper is organized as follows: In Sect. 2, recent work in detection of MAs in color fundus images is reviewed. In Sect. 3, the details of the proposed methodology for detection of MAs are presented. In Sect. 4, the results and analysis are presented. Conclusion and future scope are discussed.

2 Literature Review

Much of the related work on detection of MAs in DFP involves the following steps in common: The fundus images are first preprocessed to obtain better quality of the image and to highlight the necessary features in the image. Following preprocessing, the coarse segmentation of the fundus images is done to extract the candidate MAs. Subsequently, features are extracted from the candidate regions to distinguish false positives from true MAs. This step, called false positive elimination, is typically performed using a supervised classifier. This work addresses the coarse segmentation of MAs and hence review on literature relevant to this topic is presented below.

Spencer et al. [1], adopted subtractive shade correction and normalization for preprocessing fluorescein angiogram (FA) fundus images. Bilinear top-hat transformation was used to segment regions similar to MAs and Gaussian matched filter was employed to enhance them. Recursive region growing technique was used to extract the candidate MAs. This scheme had the disadvantage of not detecting low contrast and small MAs that were difficult to distinguish from the background. Moreover, those MAs that were conglomerated were also rejected.

In Walter et al. [6], the preprocessing was carried out on a green channel image, which provides high contrast background for dark lesions. Subtractive shade correction was carried out to alleviate non-uniform illumination in the image. Candidate MAs were detected by means of diameter closing and thresholding. In this study, an image set of 115 uncompressed digital images acquired after pupil dilation were considered were considered. The images are of size 640 × 480 with circular ROI. The major drawback in both [1] and [6] is the use of subtractive shade correction which resulted in degradation of images due to incorrect background approximation.

In Zhang et al. [7], an algorithm based on multi-scale correlation filtering and dynamic thresholding was done to extract MAs. The algorithm was evaluated on two databases namely ROC and DIARETDB1. In coarse segmentation, Gaussian kernels of different standard deviation (σ) were chosen to extract the ROIs. This was followed by adaptive thresholding to detect and eliminate the blood vessels. Higher σ value de-emphasizes sharp gradient changes in the image, thus making it more blurry.

Antal and Hajdu [8] employed dynamic selection of optimal combination of preprocessing steps and candidate extractor. Five preprocessing methods and five candidate extraction techniques were considered resulting in 25 combinations. The optimal selection of ensemble involved individual pairs to be evaluated and the final MAs were the fusion of MAs of each pair building up the optimal ensemble. Performance evaluation was tested on 199 images from three different databases namely ROC (Retinopathy Online Challenge), DIARET2.1 database and the database from Moorefields Eye Hospital, London, UK. The algorithm provided low false positive rate and low false negative rate with the use of optimal combinations, but with increased complexity and computational time taken for the system. Usage of combinational methods improves detection but with increased computational complexity.

In Zhang et al. [9], multi scale Gaussian correlation filtering (MSCF) followed by adaptive thresholding was used to locate all MA candidates. Region growing was performed on the extracted MAs and the resultant regions that were of size greater than

120 pixels were rejected. The algorithm for candidate MA detection was evaluated on the database ROC. MSCF involved the use of five different Gaussian kernels for matching MAs of various sizes. The coarse segmentation stage suffers from the disadvantage of having different scale selection which is not done automatically and could result in inaccurate detection of MAs. Increasing the number of Gaussian kernels further increases the complexity of the system.

Lazar and Hajdu [10] performed green channel extraction followed by local maximum region extraction by grayscale morphological reconstruction through breadth-first algorithm. Cross-sectional scanning was done on the image using larger cross sections of line operators. The method was tested on the ROC database. Elimination of optic disc and vessel bifurcation have not been addressed in this paper leading to false positives in the optic disc.

In Sopharak et al. [5], preprocessing was done on green channel image and denoising was done using median filter, followed by subtractive shade correction using averaging filter and contrast enhancement using contrast-limited adaptive histogram equalization (CLAHE). Then, detection and elimination of exudates and vessels were performed. Coarse segmentation of MAs was performed by using extended minima transform and diameter closing. This algorithm was also adopted by Aishwarya et al. [11] and validated on DIARETDB1 database. Subtractive shade correction resulted in incorrect background approximation. Other demerits of this algorithm were its inability to detect too small MAs and those MAs that were located near to the blood vasculature. Faint vasculatures were also left undetected.

In Tavakoli et al. [12], top-hat transform was used to decrease background variation. In order to remove small MA-like noise, averaging was done. The preprocessed image was then subdivided into several subimages. The vascular tree was then detected and eliminated by using Radon transform in all the subimages obtained, resulting in coarse segmentation of MAs. Performance evaluation was done on three different retinal image databases, the Mashhad database with 120 FA images, a local database with 50 FA images and ROC (Retinopathy Online Challenge) with 22 images. Some MAs that were located near to each other and too big MAs were wrongly detected as blood vessels.

Rosas-Romero et al. [4], computed the ratio of green to red channel for shade correction. This was followed by median filtering for denoising and pointwise pixel transformation for spatial normalization. The ROIs were extracted using bottom-hat transformation and thresholding techniques which are also adopted in [13]. This was followed by hit or miss transformation to segment the MA candidates. Too small MAs that were close to each other and conglomerated MAs were found to get eliminated in the hit or miss transformation stage. Faint MAs were left undetected due to low contrast image.

The proposed method involves the use of a simple yet robust method namely for accurate extraction of optic disc and blood vessels simultaneously, in a single step. This is done by employing bottom-hat transformation which extracts only dark regions and also performs optic disc elimination at the same time, resulting in improvement of processing speed and reduced complexity [4]. Further, it alleviates false positives resulting from improper segmentation of optic disc. Shade correction using green to red channel ratio was done as a replacement to background approximation, resulting in better image quality. In almost all related work, MAs that are close to each other and to the blood vessels were not detected properly. This paper aims at improving the detection of MA candidates through the use of modified contrast enhancement technique using morphological operations, and multiple SEs in the coarse segmentation stage.

3 Methodology

The overall flow of the proposed method is illustrated in Fig. 1. Broadly the steps involved include preprocessing, candidate extraction for dark object filtering and finally segmentation of candidate MAs.

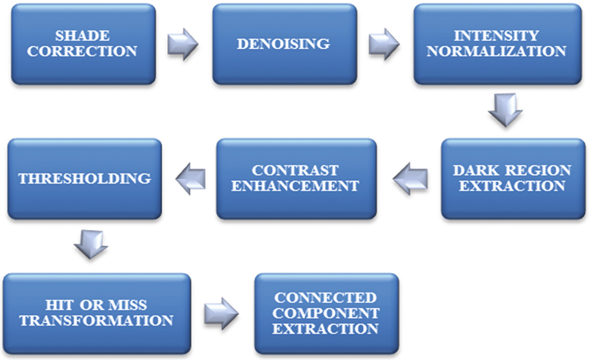


Fig. 1. Framework for coarse MA detection from color fundus images

3.1 Image Preprocessing

3.1.1 Shade Correction

The fundus images are affected by non-uniform illumination that results from factors including curvature of retina, misalignment of patient's eye and fundus camera, ocular opacities, improper dilation of pupil, poor focus of camera and inadequate illumination. This causes gradual decrease in illumination from the region of optic disc towards the periphery. The red and green channels of a fundus image contain most of the image information. The green channel of the fundus image provides the highest contrast for all blood-filled structures while red channel exhibits highest reflectance of red color and appears bright. On the contrary, the blue channel contains the least informative content as blue is absorbed by most parts of the eye. Reducing non-uniform illumination by the popular subtractive shade correction has its own demerits in choosing the appropriate size of averaging filter for background approximation. Therefore, reduction of non-uniform illumination has been performed by the red and green channels exploiting the fact that the ratio of green to red channel is a constant independent of illumination. This is computed in accordance with Eq. 1.

$$F_s\,(r, c) = [f_G\,(r,\ c)/f_R(r,\ c)] \tag{1}$$

where F_s (r, c) is the shade corrected image, f_G (r, c) is the green channel component at the row r and at column c, f_R (r, c) is the red channel component at the row r and column c.

The shade corrected image is shown in Fig. 2(b)

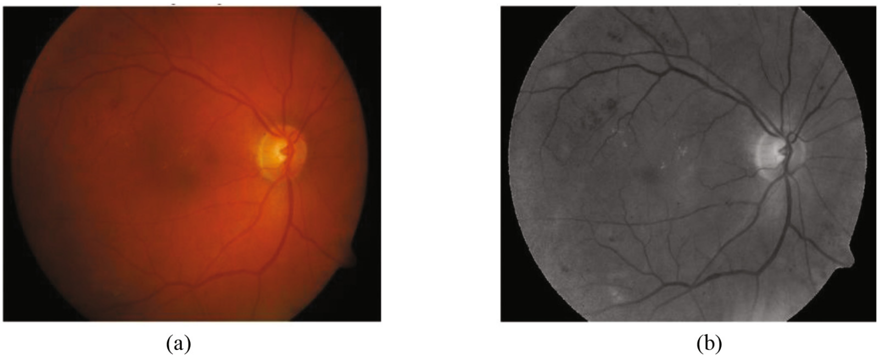


Fig. 2. (a) Original image and (b) Shade corrected image by employing green to red channel ratio

3.1.2 Denoising

The common types of noise that affect the fundus images are salt and pepper noise, shot noise and Gaussian noise. In order to eliminate the effect of noise on retinal images, the shade corrected image is denoised using a combination of median and Gaussian filter. Median filter has been proved to be effective in removal of salt and pepper noise with edge preservation while Gaussian filter provides effective noise attenuation for Gaussian noise and Poisson noise. The result of denoised image after performing shade correction is shown in Fig. 3(b)

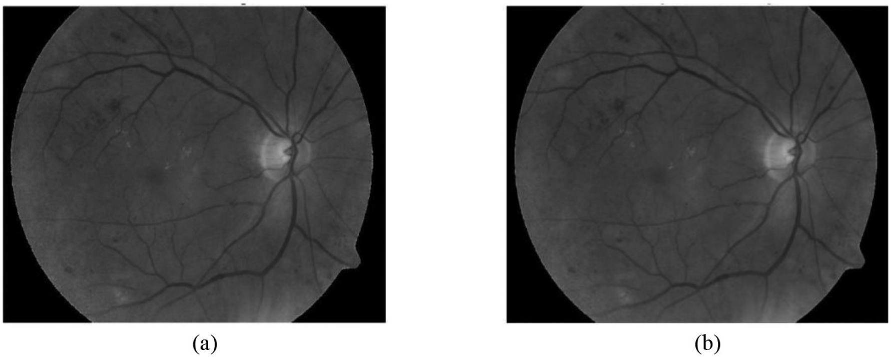


Fig. 3. (a) Shade corrected image and (b) Denoised image using median-Gaussian filter

3.1.3 Illumination Normalization

For reducing the inter-image illumination variations, which could arise due to diversity in ethnicity, illumination normalization of the image is performed using the pixel transformation using Eq. 2 [4].

$$I_a(r, c) = \frac{\sigma_n}{\sigma_I}\,(I_b(r, c) - \mu_I + 2\sigma_I) + \mu_n - 2\sigma_n \tag{2}$$

where $I_a(r, c)$ and $I_b(r, c)$ are the image grayscale values at position (r, c) after and before normalization, σ_I is the standard deviation of the image, σ_n is the reference standard deviation, median of standard deviation of all images,μ_I is the mean of the image, μ_n is the reference mean, median of mean of all images.

In this step, the mean and standard deviation of all the images get transformed to the reference mean and standard deviation value. Normalization of grayscale content plays an important role during thresholding. Proper normalization helps in choosing a single threshold value for all images. The images before and after normalization are shown in Fig. 4.

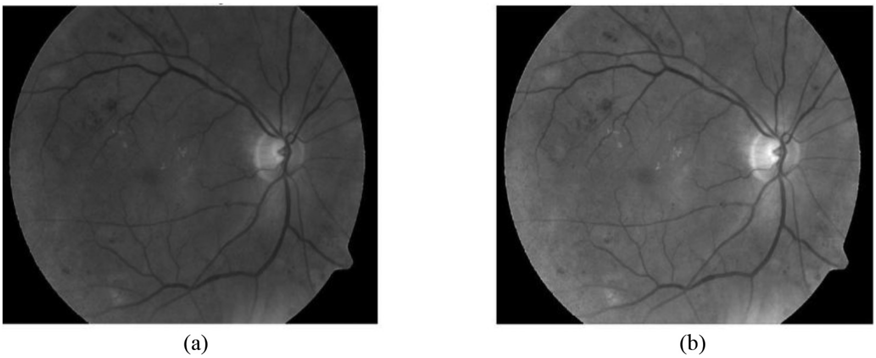


Fig. 4. (a) Denoised image and (b) Normalized image

3.2 *Bottom*-Hat Transformation

The algorithm utilizes morphological techniques to perform dark region extraction. The dark regions present in the retinal images are MAs, blood vessels, HEM and noise. The goal of the first step is that the red regions corresponding to the local minima of original image should be enhanced, while the bright regions like the optic disc corresponding to the local maxima namely should be eliminated. Bottom-hat transformation otherwise called black top-hat has been used in the proposed method for extraction of dark regions. Bottom-hat operation involves subtraction of input image from the morphologically closed image.

$$f_{bh} = [(f \bullet b) - f] \tag{3}$$

where f is the input image to this stage, b is the structuring element used for closing operation, • is the closing operator, f_{bh} is the bottom-hat transformed image.

The closing operation is performed using an SE of disk size '9', which is chosen empirically as the appropriate size of MAs. Choosing the appropriate size of SE is important in extracting all MAs. Performing two-dimensional bottom-hat transform results in isolation of certain regions of blood vessels which may be misclassified as MAs. In order to reduce the occurrence of false positives, 1D bottom-hat operation is performed row-wise and column-wise, the results of which will be combined using logical AND operation in the final step. The results of bottom-hat transformation are presented in Fig. 5(a–c)

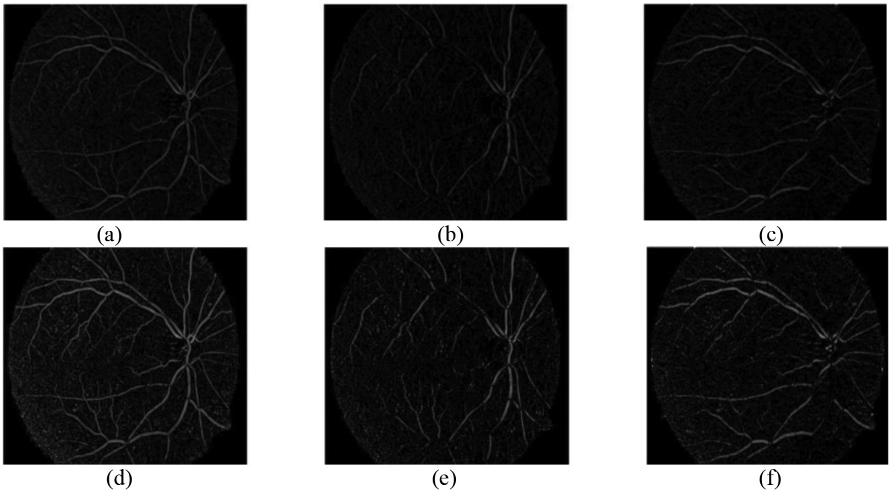


Fig. 5. (a) Results of 2D bottom-hat, (b) 1D bottom-hat over every column and (c) 1D bottom-hat over every row, (d–f) Results of contrast enhanced image for corresponding 2D and 1D bottom-hat results using morphological enhancement.

3.3 Contrast Enhancement

To further enhance the faint MAs, contrast enhancement is performed using morphological techniques. CLAHE proves to be inefficient since it does not pick up many MAs which are of low contrast. To improve the sensitivity, a combination of top-hat and bottom-hat transform is used for enhancement which retains almost all MAs in the thresholding stage. Enhancement is performed using the expression in Eq. 4.

$$A_E = A + A_{TH} - A_{BH} \tag{4}$$

where A is the input image (result of previous processing step), A_{TH} is the top-hat transformed image and A_{BH} is the bottom-hat transformed image (result of employing different size of SE), A_E is the enhanced image.

This technique improves the contrast selectively for dark lesions and blood vessels for an SE of size chosen to be 9 and 20 pixels.

3.4 Thresholding

Following enhancement, thresholding is performed. Binarization is applied to both 2D and 1D bottom-hat results using Otsu's method. Empirical choice of threshold value being employed is avoided using Otsu's thresholding. The results of thresholding operation on these outputs are presented in Fig. 6(d−f), respectively.

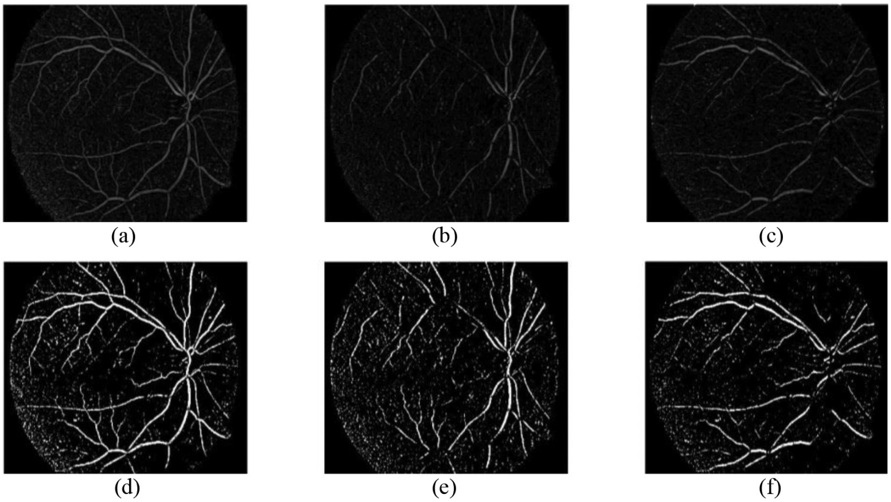


Fig. 6. (a-c) Results of enhanced bottom-hat transformed images and (d-f) corresponding thresholded images.

3.5 Hit or Miss Transformation

The thresholded image contains all the dark candidates that include MAs, blood vessels and HEMs. Appropriate methods have to be used to detect only the candidate MAs and eliminate other non-MA structures. This two-step procedure is performed using a single technique, hit or miss transformation on both 2D and 1D images separately. Hit or miss transformation is a morphological technique that can extract specific shapes of interest. Blood vessel removal is automatically achieved by the hit or miss transform, due to its ability to discriminate structures based on shape. Though HEMs are also roughly circular, they are much larger and the size-based discrimination of hit-or miss transformation is capable of eliminating HEMs in the detection process. The proposed algorithm uses SEs to exactly match the size of MAs while removing other non-MA structures and noise simultaneously, resulting in detection of MA candidates alone. Circular SEs are built using inner and outer disk structures separated by a black ground with a small white region located inside the inner disk as shown in Fig. 8. The size of inner white region is limited to a radius of 3 pixels so as to discard regions smaller than this, which correspond to noise. Don't care condition is introduced inside the inner disk for providing flexibility to match all MAs with varying sizes and irregular shapes and also outside the outer disk to detect other neighbourhood MAs. The black background helps in removal of blood vessels and in detection of two or more MAs as separate structures. Unlike MAs, blood vessels do not have the black background and thus gets eliminated in the process. MAs near to each other and close to the vasculature often get missed out. Those MAs which are clubbed together and are of size larger than the SE are also not detected. Therefore, selection of single SE cannot detect those MAs which are of large and small sizes when compared to the typical size and also those which are clubbed together or overlapped with each other.

To improve sensitivity of MA detection, SEs of different sizes are chosen to accommodate all possible MA candidates that do not fit in the particular size of SE. MAs are irregular shaped structures that are approximately 9 pixels in size. Thus the optimum size of circular SE is chosen with an inner radius of 9 pixels and outer radius of 11 pixels with a 2 pixel gap for the black background. Thus the lower limit on the detected regions of 3 pixels is imposed by the white region while the don't care region imposes an upper limit on the inner radius of up to 9 pixels. Similarly, other sizes of SEs are chosen with inner and outer radii of 6 and 7 pixels, 13 and 15 pixels, 18 and 20 pixels respectively. Smaller SE with 6 ad 7 pixels radii is chosen to detect smaller MAs and those that were partially detected in the binarization stage, also retaining MAs that are close to each other providing a gap of 1 pixel. The large radii SE of 13 and 15 pixels are used to detect larger MAs, and those MAs that were clubbed together in thresholded image are detected using SE of 18 and 20 pixels radii. By adopting various SEs, a significant increase in the detection results was achieved. The optimum size SE (9 and 11 radii) is shown in Fig. 7. The corresponding candidate MA extracted is shown in Fig. 8. The resulting images of different SEs are combined using logical OR operation for their respective 2D and 1D images in the latter stage.

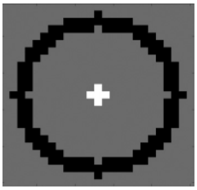


Fig. 7. Optimum size of SE

3.6 Extraction of Connected Components

The extracted candidate MAs do not cover the entire region encompassed by the MA. To recover the entire shape of MA and to eliminate those regions of blood vessels that are still detected in the hit or miss transformation stage, extraction of connected

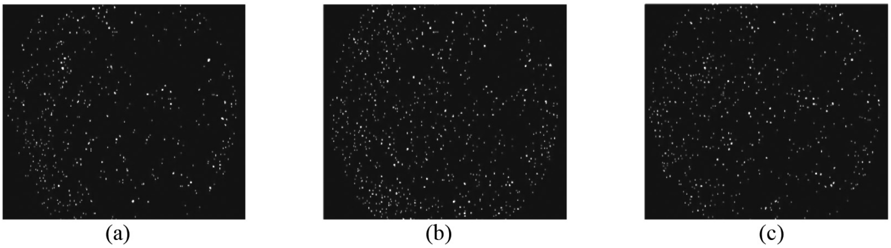


Fig. 8. (a) Results of hit or miss transformation with 2D thresholded image, (b-c) with 1D thresholded image.

components as in Eq. 5 is performed with 3 iterations on the binary image resulting from hit or miss transform.

$$X_K = (X_{K-1} \oplus b) \cap f \tag{5}$$

where $X_{K\text{-}1}$ is the image dilated with the structuring element 'b', until the complete shape of the component is extracted,

$\oplus$ is the dilating operator,
b is the suitable 5 × 5 square SE for performing dilation operation,
f is the thresholded image used for extraction of connected components,
X_K is the extracted component image.

The images after extraction of connected components contain portions of blood vessels along with extracted MAs which are eliminated by performing logical AND operation on 2D and 1D images. The 1D image along vertical column will recover only vertically oriented blood vessels and 1D image along every row will recover only horizontal blood vessels and 2D image extraction will result in recovering both horizontal and vertical blood vessels. Hence the common portions of blood vessels are

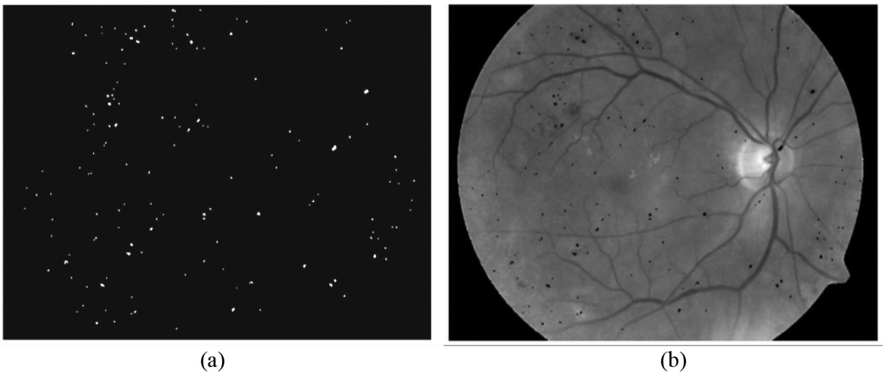


Fig. 9. (a) Candidate MAs after extraction of connected components, (b) Fundus image with the detected MAs after coarse segmentation.

Table 2. Various structuring elements used to perform hit or miss transformation

Inner & Outer Radiuses	Necessity
6 & 7	To fit in MAs that were partially detected in the binarization stage also retaining MAs that are close to each other
9 & 11	Optimum size of MAs with a gap of 2 pixels
13 & 15	To detect Large MAs
18 & 20	To detect MAs that are clubbed together and detected as single element during binarization.

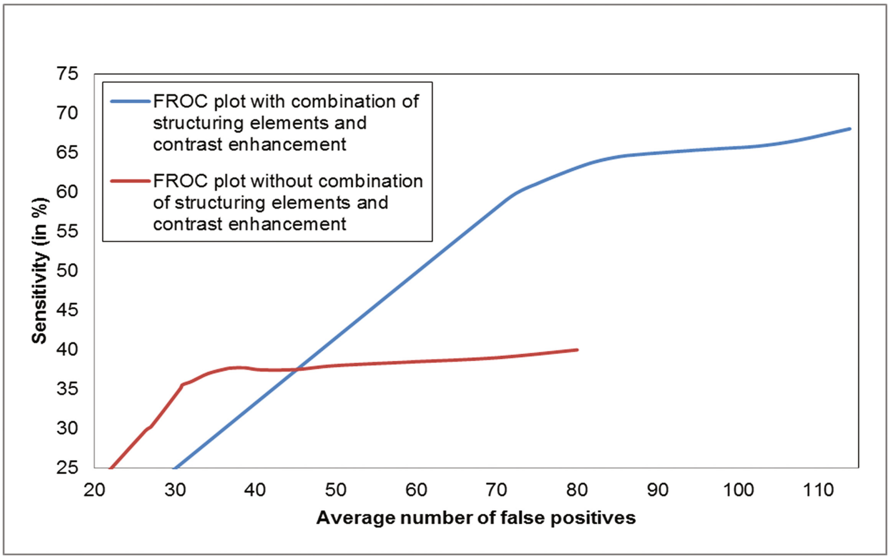


Fig. 10. FROC plots comparing the coarse segmentation results of the proposed method (curve marked in blue) with the previously used method (curve marked in red). (Color figure online)

alone extracted by performing AND operation of images, reducing the occurrence of false positives. The results of extraction of connected components after performing logical AND operation is shown in Fig. 9.

4 Results and Conclusion

4.1 Database Description

The color fundus images considered in the study were taken from DIARETDB1 database. All the images were captured with the 50° field-of-view digital fundus camera. There are totally 89 images which were taken in the Kuopio University hospital. Out of the 89 images, 84 contain at least mild non-proliferative signs (MA) of the diabetic retinopathy and 5 are considered as normal which do not contain any signs of diabetic retinopathy. Ground truth images annotated by expert groups are provided for reference.

4.2 Coarse Segmentation Results

The performance of MA detection using the proposed method is analyzed using FROC curve obtained using varying the threshold value in the binarization stage. The FROC curve plots average number of false positives to the sensitivity obtained for varying values of threshold. A set of 20 images each with approximately 30 to 50 MAs were considered. The FROC curve for the proposed method was obtained with the use of contrast enhancement through modified morphological enhancement, also employing the use of combination of all four SEs whose size and role are tabulated in Table 2. This plot was compared to the FROC curve obtained through the use of only the optimum size of SE in the hit or miss transformation stage, with no contrast enhancement. The results are shown in Fig. 10. It can be observed from the results that there is drastic up shift in the curve obtained through the proposed method, verifying that the sensitivity of MA detection has been improved. This improvement in MA detection rate is because of the ability of the proposed work to capture the difficult cases including small MAs, faint MAs, and MAs close to each other and to the vasculature, which are overloaded by the existing systems. In future, this work could be extended by incorporating fine segmentation to eliminate false positives. Increase in false positives in the attempt to increase the sensitivity of the system can therefore be addressed by the fine segmentation stage.

4.3 Conclusion

The result of coarse segmentation stage has not been reported in many papers. Only the performance of the classifier employed in the fine segmentation stage (false positive eliminated) has been discussed much. As a result, there is no true picture on the number of missed out MAs during coarse segmentation. This paper has successfully reported the results of coarse segmentation thus presenting the true sensitivity rate. From the coarse segmentation results obtained, it can be seen that the proposed method has achieved higher sensitivity by detecting almost all MAs that were difficult to detect otherwise. The contrast enhancement using modified morphological enhancement improved the detection of faint MAs that were difficult to distinguish from the background. Choosing and employing different SEs for detection of MAs greatly improved sensitivity by picking up the MAs that were getting missed out in other methods of detection process. Improving MA detection rate in the coarse segmentation is important to achieve an overall high sensitivity. Further reduction of false positives is carried out in the fine segmentation stage.

References

1. Spencer, T., Olson, J.A., McHardy, K.C., Sharp, P.F., Forrester, J.V.: An image-processing strategy for the segmentation and quantification of microaneurysms in fluorescein angiograms of the ocular fundus. Comp. Biomed Res. **29**, 284–302 (1996)
2. Jackuliak, P., Payer. J.: Osteoporosis, fractures, and diabetes. Int. J. Endocrinol. 2–10 (2014)
3. Winston, Dr., Scott, J.: Diabetic retinopathy. http://wjscottmd.com/diabetic-retinopathy. Accessed 11 Sept 2016
4. Rosas-Romero, R., Martínez-Carballido, J., Hernández-Capistrán, J., Uribe Valencia, L.J.: A method to assist in the diagnosis of early diabetic retinopathy: Image processing applied to detection of microaneurysms in fundus images. Comput. Med. Imaging Graphics **44**, 41–53 (2015)
5. Sopharak, A., Uyyanonvara, B., Barman, S.: Simple hybrid method for fine microaneurysm detection from non-dilated diabetic retinopathy retinal images. Comput. Med. Imaging Graph. **37**, 394–402 (2013)
6. Walter, T., Massin, P., Erginay, A., Ordonez, R., Jeulin, C., Klein, J.-C.: Automatic detection of microaneurysms in color fundus images. Med. Image Anal. **11**, 555–566 (2007)
7. Zhang, B., Wu, X., You, J., Li, Q., Karray, F.: Detection of microaneurysms using multi-scale correlation coefficients. Pattern Recogn. **43**, 2237–2248 (2010)
8. Antal, B., Hajdu, A.: An ensemble-based system for microaneurysm detection and diabetic retinopathy grading. IEEE Trans. Biomed. Eng. **59**, 1720–1726 (2012)
9. Zhang, B., Karray, F., Li, Q., Zhang, L.: Sparse representation classifier for microaneurysm detection and retinal blood vessel extraction. Inf. Sci. **200**, 78–90 (2012)
10. Lazar, I., Hajdu, A.: Retinal microaneurysm detection through local rotating cross- section profile analysis. IEEE Trans. Med. Imaging **32**, 400–407 (2013)
11. Aishwarya R., Vasundhara T., Ramachandran K. I.: A hybrid classifier for the detection of microaneurysms in diabetic retinal images. In: Goh, J., et al. (eds.) The 17th International Conference on Biomedical Engineering, ICBME Proceedings, vol. 61, pp. 97–103. Springer, Singapore (2017)
12. Tavakoli, M., Shahri, R.P., Pourreza, H., Mehdizadeh, A., Banaee, T., Toosi, M.H.B.: A complementary method for automated detection of microaneurysms in fluorescein angiography fundus images to assess diabetic retinopathy. Pattern Recogn. **46**, 2740–2753 (2013)
13. Devi, S.S., Ramachandran, K.I., Sharma, A.: Retinal vasculature segmentation in smartphone ophthalmoscope images. In: 7th WACBE World Congress on Bioengineering 2015, pp. 64–67. Springer International Publishing (2015)

Intelligent Recognition Techniques and Applications

Swarm Robots in a Closed Loop Visual Odometry System by Using Visible Light Communication

Dhiraj Patil, Kewal Shah, Udit Patadia, Nilay Sheth(✉),
Rahul Solanki, and Anshuman Singh

VJTI, Mumbai, India
patil.dhiraj107@gmail.com, kewal.m.shah@gmail.com,
udit7395@gmail.com, nilay_994@hotmail.com,
solankirahul411@gmail.com, anshuman786singh@gmail.com

Abstract. Motivated by the looming radio frequency spectrum crisis, this project aims at demonstrating that Visible Light Communication (VLC) has now reached a state where it can prove that it is a viable solution to this fundamental problem. VLC is a technique used for data transmission at very high speeds through light, which transfers data by varying its intensity at unperceivable rates. The proposed solution also establishes a closed loop with an overhead camera which mocks like a GPS in the swarm environment. Positional information of the robots is given by augmented reality tags and this data is fed-back to the robots via VLC. The paper elaborately describes the approach used to exploit off-the-shelf components for facilitating VLC. Overhead localization and the closed loop made here control each swarm robot with simplex communication. Problems faced while prototyping and overcoming them in revisions have been also described.

Keywords: VLC · Visual odometry · Swarm robots · Closed loop P controller · Li-Fi · Augmented reality tags · Pose estimation · Analog filtering · Dynamic path planning

1 Introduction

The primary purpose of this paper is to highlight the ease with which a Visible Light Communication (VLC) system can be prototyped to communicate in a single master multiple slave simplex asynchronous indoor robot navigation systems. Secondary purpose of the paper is to demonstrate the different algorithms that are used for coordinating swarm robots [5]. Hence, the paper aims to establish a closed loop which controls swarm robots via VLC [1].

The proposed solution in the paper proves that a 19.2 kbps system was enough to push data packets to multiple swarm robots driven by visual odometry. Also, the pros and cons of the algorithms used to implement path planning are stated. The entire system when brought together illustrated that VLC can be implemented using off the shelf components (For schematics Ref. Fig. 1). The circuit developed had a long range unlike most common VLC systems and functioned on simple baseband modulation [3].

S.M. Thampi et al. (eds.), *Advances in Signal Processing and Intelligent Recognition Systems*, Advances in Intelligent Systems and Computing 678,
DOI 10.1007/978-3-319-67934-1_18

The ruggedness and noise immunity of the VLC circuit proposed provides flawless error free wireless connection [2]. It shows that using camera for feedback and error computations at centralized system is a quick way to guide robots. This paper explores different path planning algorithms and shows how dynamic path planning is better in this scenario where indoor robotic navigation is required in real time [11].

VLC is established between the overhead camera with a LED transmitter and ambient light sensors as receivers on our swarm robots. The baseband modulation used in VLC was optimized to 19.2 kbps at a distance of 3 m. The overhead camera acts as a GPS for the robots and helps to determine the pose estimation data of the robots. Central system estimates the path and updates the robots with the relative error to the destination. This error is encapsulated in a packet and broadcasted over the workspace via visible light. This error then runs through a P controller mapped to the PWM on the differential drive motors. This visual odometry with dynamic path planning [10] was tested to establish a loosely coupled closed loop between the motors on the swarm bot and the central system which instructs the robot.

The paper is organized as follows: Sect. 2 expands on the concept of VLC and visual odometry and compares the currently used control systems with the proposed system in the paper. Sections 3 and 4 talk about the hardware and software implementation and also elaborate on other concepts like path planning and packet composition used while devising the platform. Section 5 discusses the methodology implemented in the algorithms and how they were selected to suite this swarm environment. Section 6 summarizes the parameters of the system with the results. Section 7 highlights the applications of VLC and concludes the paper.

2 Description and Related Work

Visible Light Communication: VLC exploits the bandwidth of visible spectrum for communication and hence is based on LEDs for the transfer of data. Variation in the intensity of light provides us binary 1 and binary 0 of data which transmits information wirelessly in case of baseband modulation. Freeing up the RF spectrum density and speeding up data transmission is one of the major purposes of VLC.

On comparing the most commonly developed VLC systems, they are short range [4]. Modulation techniques like OFDM, OPSK, QPSK, FSK, etc. make the VLC receivers and transmitters too complex since they involve DSP processors and require DFT (Discrete Fourier Transform) capabilities. We implement simple baseband communication. Shortcomings in many systems are due to peak wavelength mismatch between phosphor based LEDs and the peak sensitivity of the receivers [4]. Most commonly used VLC trans-receivers use straight line communication, which cannot be used in dynamic robots. The receiver design proposed in the paper can receive light in a solid cone of 60° with matching peak sensitivities.

Closed Loop Visual Odometry: Odometry is the use of data from motion sensors to estimate change in position over time, hence an important aspect in robotics [10]. The robots need closed loop control to go from one position to other since open loop techniques are largely inaccurate while doing so. Common methods involve using

LIDAR (expensive) and wheel encoders (slips and errors). However in this project, we have used an off-board camera to monitor the robot to guide it through a desired path. Hence collectively called closed loop visual odometry.

Commonly, robots use encoders, LIDAR, SLAM and proximity sensors to carry out odometry [5]. These techniques are computationally heavy and need onboard sensors. The proposed solution can work on robots having no sense of surroundings by using a single overhead camera.

The common controllers are generally closely coupled i.e. the error calculations are handled on the same microcontroller which is also responsible for taking inputs (encoders in closely coupled systems) and mapping outputs. This paper explores the ability of establishing a closed loop remotely (loosely coupled), where the computer calculates the errors after path planning and sends it wirelessly to the robots.

Table 1 discusses the differences on approach used in implementation of the objective of this paper as compared to the current widely used approaches in odometry, controllers and path planning.

Table 1. Difference from majority implementations on control systems

Sr. No.	Differences from current control systems		
	Aspects	Current common control systems	Proposed system in the paper
1.	Odometry	Physical [5]	Visual
2.	Controllers	PID closed loop Tightly Coupled	P closed loop Loosely coupled
3.	Path Planning	Static and Computationally Heavy[11]	Dynamic and Lite
4.	Communication Bandwidth	Limited RF B.W. [1]	10 times wide, visible spectrum

3 Hardware Implementation

A. ***VLC Receiver***

VLC receivers used here were high speed ambient light sensors. The function of the receiver was to handle the following [6, 9]: *Amplification, noise-filtering, thresholding, No loading effects on any intermediate stage and had to be compatible with microcontroller's noise margin levels.* TEMD6200 [8] was chosen as the VLC receiver since it supported higher baud rates, due to low junction capacitance. It received any light with a peak sensitivity in the green color wavelengths of 540 nm.

Stages of Receiver: Fig. 1 below describes the cascade implementation using Analog filtering Op-Amps [7].

TEMD6200 > Trans-Impedance > Buffered Notch > Bandpass Filter > Comparator > Buffer > Swarm MCU

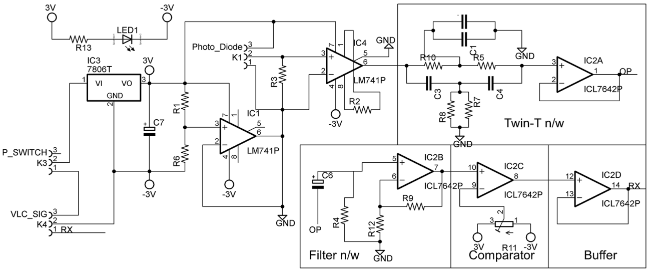


Fig. 1. Schematics for 19.2 kbps VLC receiver

B. ***VLC Transmitter***

VLC transmitter used is a microcontroller based switching device which modulates the light at high baud rates as instructed by the central system. Apart from just modulating, the transmitter stores the last updated packet from the central system and keeps broadcasting it continuously until a new packet arrives. This way, the link is continuously occupied and corruption of data due to random signals is avoided. If the receiver is not flooded with packets, it might accept noise signals of the link which leaves garbage data in the receiver buffer. The paper implements a phosphor LED based VLC transmitter.

C. ***Swarm Robot design***

Setting up the environment involved integrating all the modules together to finally achieve the exact goal of the paper. Camera and transmitting LED with transmitter microcontroller were mounted overhead, 1 m above the workspace. The microcontrollers were flashed with the firmware and the motors were mounted at exact angles for equal traction on both ends. One shield powered the microcontroller with the motor driver and the other shield powered the VLC receiver circuit *(Ref.* Figure 2).

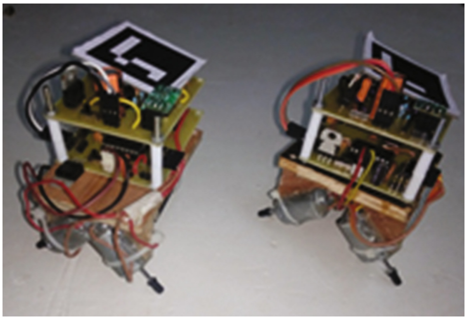


Fig. 2. Swarm robots in the workspace

4 Software Implementation

A. *Communication Protocol: Packets and Parsing*

Packets were based on UART protocol (Ref. Fig. 3). They were simple bit-strings of the ASCII values of the characters with start and stop bits but without flow control and are sent from the central system via VLC. The transmitter end makes sure that the forward channel is occupied with the last updated packet from the central system. This is done by repeating the latest packet on the transmitter node, unless a new packet arrives. Figure 4 shows the frames which make up a packet. Header frame notifies the system as start of packet. Packet ends with Trailer frame appended with *“\n”* (newline constant). The Instruction frame updates the robots if they should be in dynamic mode or static mode. The unique robot ID makes sure that the packet gets associated to a particular robot.

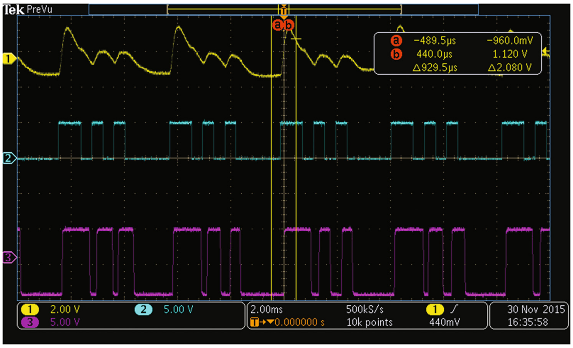


Fig. 3. Oscilloscope Readings for our VLC circuits. Trace Legend: ch. 1- Yellow Trace- Ambient Light sensor Raw. ch. 2- Blue trace- VLC Transmitter. ch. 3- Purple Trace- VLC Receiver (Color figure online)

Frame	***Header***	***Instruction***	***Robot ID***	***Error***	***Trailer***
Length	5 bytes	2 bytes	3 bytes	3 bytes	4 bytes
Description	“start” Start of frame	“30” : go “25” : stop	ArUco robot ID	Heading error offset by 180	“stop” End of frame
Example	Start	30	768	200	stop

Fig. 4. Packet Composition

Example frame decoded: “start 30 768 200 stop”: Indicates that robot no. **“768”** should be in “**Go**” mode with a heading error of “**200”** − 180 = 20.

Robot updates this information as follows:

pwm_right = fwd_speed +20;

pwm_left = fwd_speed −20;

If fwd_speed = 80, right motor will go forward with a speed of 100 and left motor will go forward with a speed of 60, then robot will differentially take a left.

Formula for Refresh rate for Robot's path correction

(i) The robot orientation correction value is carried in the frame which is 3 bytes long.
(ii) The length of the entire packet is 5 + 2 + 3 + 3 + 4 = 17 bytes (Fig. 4.)
(iii) The baud rate currently is 19.2kbps = 2.4 kB/sec
(iv) The frames are sent out sequentially to N robots in the workspace.
(v) Packet update rate will hence be 2400/(17 × N).
(vi) For N = 3, Robots are updated @ 47 Hz
(vii) Resultant refresh rate(R) is the convolution of the two frequencies; namely
(viii) Pose estimation rate (fps ≈30 Hz)……. (*Ref.* Table 4)
Error correct transmit baud rate (f_c ≈47 Hz)
(ix) It can be given by $f_c \parallel \text{fps} \approx 18.31$ Hz

As seen above, the refresh rate is confined due to visual pose estimation algorithm. With this in place, the robots could now navigate from A to B, by parsing these packets and feeding the loosely coupled error to the P controller to establish a closed loop. However this *unicast packet* had to be polled in round robin and generate errors for each robot ID to move multiple robots at a time. To shorten the sequential packet, a single *broadcast packet* can be composed which contains heading errors of all the robots to drive them concurrently.

B. *Path Planning*

After sending the robots in a straight line from A to B, a path planning approach was required to prevent collisions between the moving robots in the workspace. The paper discusses three algorithms, a comparison of which is given in Table 2.

i. *A Star Algorithm*: Traditional AI approaches involve using the A* like majority of computer games use [11]. For e.g. in strategic games, army is guided from A to B so they find a minimum path to the attack location. However A* requires heavy pre-processing and the workspace must be discretized with cost functions.
ii. *Diversion Node Algorithm*: A very lite approach can be used to avoid collisions, by simply redirecting the robot to a perpendicular direction called as the diversion node. However it does not ensure a minimum path to goal.
iii. *Force Field Algorithm*: To eliminate the above problems, we implemented real time force field calculations, where various attraction and repulsion vectors guide the robot from A to B. The algorithm can be explained mathematically by using the following terms and graphically by Fig. 5 [11].

Q_r: Robot Potential
C: Target Potential
r_t: Vector from target to Robot
B_i: Barrier term
K: No. of robots in the workspace

L: No. of boundaries

r_b: Perpendicular vector from boundary to robot

$$\mathrm{F} = Q_r\left[C\frac{r_t}{\|r_t\|^2} + Q_r\sum\nolimits_{i=1}^{K-1}\frac{r_i}{\|r_i\|^2} + \sum\nolimits_{i=1}^{L} B_i\right] [11]$$

$$where,\ B_i = B_r\left(\frac{r_b}{\|r_b\|^2}\right)$$

C term is the attraction vector to target, while Q_r and B_r are repulsions from obstacles and Boundaries respectively. The B_r term behaves differently than the Q_r term, by guiding the robot radially away from the boundary tangent. However C term and Q_r terms are the attraction and repulsion vectors respectively, inversely proportional to distance.

Figure 5 is generated by plotting the above equation graphically. Assume the obstacle is located at (200, 200) and target is located at (128, 128). The skewed plane of attraction extrudes at (128, 128) and (200, 200) indicating maximum attraction and repulsion points respectively. Any robot placed in the force field at any co-ordinate is forced to be absorbed by the blue region and pushed away from the red region to minimize the force field potential. After reaching the desired target, the force field converges to zero and the robot becomes stationary.

Table 2. Algorithms for path planning

Sr. No.	Algorithms tested on the Swarm Robots for path planning		
	Algorithm used	Advantages	Disadvantages
1.	A*	1. Heuristic analysis and solution	1. Works only on Static frame 2. CPU Intensive 3. Discrete plane
2.	Diversion Node	1. No load on CPU 2. Dynamic frame 3. Continuous plane	1.Obstacle avoidance not guaranteed 2. Not the shortest path to goal
3.	Real time force field calculations	1. Not very CPU intensive 2. Dynamic frame 3. Continuous plane	1. Not the shortest path

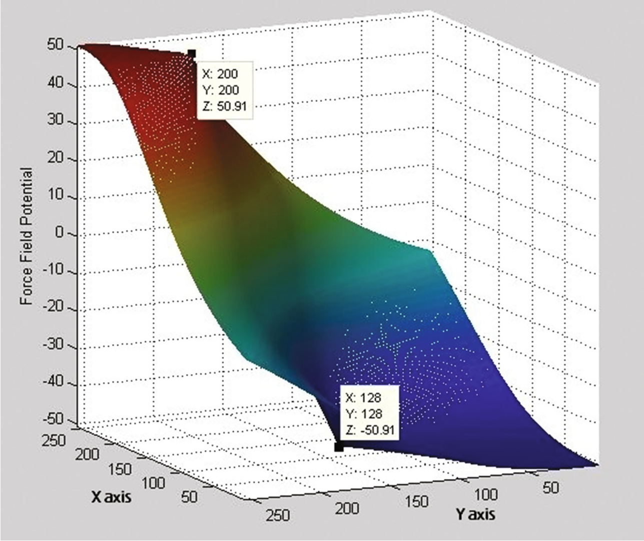


Fig. 5. Force field distribution in the workspace

C. *Algorithm Implementation*

Software complied had to take the following responsibilities: Taking input from the camera, computing the error for each robot and broadcasting the packet which translates into VLC data [12]. It was mandatory for the software to maintain a high FPS rate to keep up with the correction rates on the robot. To solve this purpose, the software uses Open Computer Vision library which is compiled on C++. The laptop (central system) running the software is connected with a camera for capturing odometry information. This odometry information is used to calculate the heading of the robots and this unique data is encoded in packets. Packets are forwarded to the MCU which is responsible for modulating VLC.

The **first** algorithm imposed a single point constraint on the tracking algorithm. In a single point constraint, one or more degrees of freedom of the robot can be interpreted for a given point in space [11]. Due to this in some cases, namely when robot acquires instructions for obtuse angled corrections, it lost its target.

This was solved by the **second** algorithm which applied Camshift algorithm for robot's contours and an additional infrared LED on the robot which gave information about its orientation. After broadcasting the packets to all robots, the one with an identity match would indicate its availability by turning on its infrared LED. However this method was CPU intensive and inefficient.

To speed up the software and avoid intensive calculations, in the **third** algorithm, an augmented reality library was included which used ArUco markers [13] for pose estimation. Each marker could have a unique ID. This eliminated the need of an infrared LED. It also facilitated a multiple dynamic robot workspace at any instant i.e. all robots can move simultaneously. In the second algorithm, they had to wait for other robots to stop and turn off their respective LEDs and then one of the robots could traverse.

The packets in the third algorithm were modified to deliver concurrent instructions at a uniform data rate instead of a round robin polling sequentially to update the robots. The communication medium still being visible light, the robots would stop if a strong opaque shadow covers it or even stop when they tried to escape the workspace where light couldn't reach.

5 Methodology

(i) Overhead camera sends live feed to the image processing platform which acts as a central system.
(ii) The tracking algorithm running on the central system updates the local co-ordinate of each robot.
(iii) Orientation errors with respect to destination are calculated and the error angle is updated.
(iv) Packets (Fig. 5) are composed with the swarm ID and the corresponding error.
(v) Packet strings are sent out serially to VLC transmitting microcontroller. It sends out the last updated packet continuously to VLC receivers of the swarm robots.
(vi) The packet is decoded after processing on swarm robots.
(vii) The errors are mapped directly via P controllers to the motors. The robot refreshes motor speeds and the entire process repeats.

Tables 3 and 4 give in depth details about the three algorithms implemented based on FPS, tracking and concurrent control of swarm robots.

Table 3. Methodology of tested algorithms

Sr. No.	Methodology of Algorithms for tracking Swarm Robots	
	Algorithm	Steps in detail
1.	HSV tracking	1.Threshold via HSV 2.Convert to binary image 3.Erode and Dilate to remove noise 4.findContours 5.Moments of contour for Centre of Mass 6.Plot lines and select goal points
2.	Camshift	1.Back Projection with histogram normalization 2.meanShift to cluster contour 3.Track contour 4.Plot lines and select goal points
3.	ArUco [13]	1.Adaptive Thresholding 2.findContours 3.Polygonal approximation of 4 corners 4.Refine corners using subpixel interpolation 5.Frontal perspective of marker by homography 6.Otsu thresholding and hamming decode 7.Plot lines and select goal points

Table 4. Features of tracking algorithms for swarm robots

Sr. No.	Features of Algorithms for tracking Swarm Robots			
	Aspect	1st Algorithm	2nd Algorithm	3rd Algorithm
1.	Detection	HSV threshold	Camshift	ArUco
2.	Robot Orientation Info	Not possible	Given via IR Led and Centre of contour	Possible via Pose estimation
3.	Process Threading	Single thread CPU	Multithread GPU	Single Thread CPU
4.	FPS	15	38	28
5.	Efficiency	Least	Better	Best
6.	Robot ID	Not possible	IR Led glows	ArUco ID
7.	Multirobot control	Not possible	Complex	Simple
8.	Packet	Sequential Round Robin	Sequential Unicast	Concurrent Broadcast
9.	Error Calc.	Single point constraint	Heading error is the error	Heading error is the error

6 Results

The advantage of having a visual odometry system is that it acts as a local GPS while also providing orientation data without any sensor fusion, using the camera alone. The paper also shows that this camera feed can be used to establish a loosely coupled closed loop P controller to guide robots from A to B. The purpose of using VLC instead of RF was to explore the scope of receivers in a workspace capable of filtering VLC data and adjusting to any ambient lights in the environment. VLC served as a robust simplex asynchronous communication medium in the hemisphere [3]. The circuits being cheaper than RF techniques, VLC was dependable enough to ease up the radio spectrum where simplex asynchronous data at 19.2 kbps was sent. Table 5 sums up the salient features of the three major aspects covered in the paper:-

i. Feasibility of loosely coupled visual odometry.
ii. VLC can be a strong alternative to RF, even in broadcast/unicast dynamic environments.
iii. Force Field path planning is an efficient algorithm in indoor navigation systems.

Table 5. Features of the Odometry system on VLC

Sr. No.	Closed loop VLC Odometry parameters	
	Parameters	Value
1.	Baud Rate	19200 bps
2.	Range	3 m, 60 degree solid angle
3.	Controller	Loosely coupled P controller
4.	FPS	28 on Single thread Intel i5 @2.8 GHz
5.	No. of Robots in system	3
6.	Path planning and Detection Algorithm	Dynamic Force Field distribution on ArUco Algorithm [13]

7 Conclusion and Future Scope

Firstly, the paper successfully explores the scope of VLC communications by demonstrating that swarm systems could be easily co-ordinated by a single light source while also establishing a closed loop over the dynamics of the robot. Off the shelf components used while designing a 19.2 kbps VLC system demonstrate that the project can be replicated and scaled easily. Also, by using this technology more prominently, Wi-Fi can be less cluttered and VLC bandwidths can be explored for nominal data transfers.

Secondly, the ease with which the augmented reality markers were used as compared to our initial method of tracking with Camshift was explored. ArUco is widely used for real time augmented reality based pose estimation.

Thirdly, the feasibility of establishing a closed loop system on the swarm robots with motors without physical feedback was explored using our tracking algorithm. The mock-GPS recognized the ArUco markers at a coherent FPS while motors on the robots were updated almost without any latency.

Applications can range from indoor localization to headlights on self-driving cars. Also VLC can be used in hazardous areas like gas stations and can be used for underwater communications. LEDs will be not only use as a source of illumination but also as a medium for wireless indoor communication. We believe that the gaining notoriety of VLC/Li-Fi can not only give a relief to the RF spectrum but also be a cheaper and quicker way to deliver information.

Acknowledgment. We would like to thank Dr. Faruk Kazi, CoE-CNDS and Society of Robotics and Automation, VJTI for providing the apparatus and instruments so as to create this odometry system for swarm robots. We would also like to extend our gratitude to Prof. Harald Haas, University of Edinburgh to inspire us to model this VLC system. The motivation to innovate was inculcated by ABU Robocon, an International Competition for enriching diverse knowledge in the field of Robotics. We are grateful to Texas Instruments for providing necessary hardware under their university program which helped us learn about embedded systems and filter applications.

References

1. Haas, H., Yin, L., Wang, Y., Chen, C.: What is LiFi? J. Lightwave Technol. **34**(6), 1533–1544 (2016)
2. Mangold, S.:, Using consumer LED light bulbs for low-cost visible light communication systems. In: Workshop on Visible Light Communication Systems (VLCS) at Maui, Hawaii, pp. 9–14, September 2014
3. Elgala, H., Mesleh, R., Haas, H.: Indoor optical wireless communication: potential and state-of-the-art. IEEE Commun. Mag. **49**(9), 56–62 (2011)
4. Mangold, S.: LED-to-LED visible light communication networks. ACM International Symposium on Mobile Ad Hoc Networking and Computing (ACM MobiHoc) at Bangalore, India, pp. 1–10, August 2013
5. Rubenstein, M., Ahler, C., Nagpal, R.: Kilobot: a low cost scalable robot system for collective behaviors. In: 2012 IEEE International Conference on Robotics and Automation (ICRA). Saint Paul, MN, pp. 3293–3298 (2012)
6. Orozco, L.: Application note on "Optimizing Precision Photodiode Sensor Circuit Design." In: Analog devices application note MS 2624, August 2014
7. Carter, B.: Application note on "Filter Design in Thirty Seconds". In: High performance analog texas instruments. ti.com/lit/an/sloa093/sloa093.pdf. Accessed Mar 2015
8. Datasheet: TEMD6200FX01, Vishay semiconductors. vishay.com/docs/81812/temd6200.pdf. Accessed Aug 2015
9. Vishay semiconductors: Circuit and Window Design on filtering, amplifying and window size of Ambient Sensors. vishay.com/docs/84154/appnotesensors.pdf. Accessed Jun 2015
10. Jain, P.: Odometry and motion planning for omni drive robots. In: Computational Intelligence on Power, Energy and Controls with Their Impact on Humanity (CIPECH), Ghaziabad, pp. 164–168 (2014)
11. Feng, B., Gao, Y.: Development of strategy software algorithm simulators for multi-agent system in dynamic environments technology and communication. VAMK, University of Applied Sciences (2013)
12. Tippenhauer, N.O., Giustiniano, D., Mangold, S.: Toys communicating with LEDs: Enabling toy cars interaction. In: 2012 IEEE Consumer Communications and Networking Conference (CCNC), Las Vegas, NV, pp. 48–49 (2012)
13. Boby, R.A, Saha, S.K.: Single image based camera calibration and pose estimation of the end-effector of a robot. In: 2016 IEEE International Conference on Robotics and Automation (ICRA), Stockholm, pp. 2435–2440 (2016)

An Adaptive Neuro-Fuzzy Inference System Based Situation Awareness Assessment in VLC Enabled Connected Cars

P. Balakrishnan[1(✉)], Gnana Guru Ganesan[2],
Ezhilarasi Rajapackiyam[2], and Umamakeswari Arumugam[2]

[1] SCOPE, VIT University, Vellore Campus, Vellore, India
baskrish1977@gmail.com
[2] School of Computing, SASTRA University, Thanjavur, India

Abstract. Intelligent Transportation Systems (ITS) demand driving safety as an eminent design requirement for future generation vehicles. Collision evasion as well as consequent casualties minimization command timely delivery of significant precautionary information to the drivers. Consequently, the driver may get a clear view about the present driving situation and be able to adopt timely decision to circumvent the forthcoming dangers. This research work proposes an Adaptive Neuro-Fuzzy Inference System (ANFIS) based situation assessment method that supports the drivers to take up suitable decisions by analyzing the driver behavior of preceding/succeeding cars. The proposed approach models the stability of drivers in the perspective of connected cars and deduce the current stability situations from the sensors which are implanted in the cars. This connected cars scenario for Collision Warning System (CWS) is simulated using three Raspberry Pi boards along with ultrasonic sensor, gas sensor and accelerometer sensor. These sensor data are transmitted to other preceding or succeeding cars using visible light communication. Subsequently, these data are processed using both Mamdani and ANFIS model for situation assessment which provides the stability level of drivers. The result concludes though the Mamdani model quickly computes the stability of driver by analyzing the sensor data, it suffers from low sensitivity and precision when compared to ANFIS which showcases higher sensitivity and precision.

Keywords: Connected cars · Visible light communication (VLC) · Mamdani fuzzy logic and ANFIS

1 Introduction

Smart cities are a vision for urban development that intelligently manages the city infrastructure thereby enhancing the quality of life. Smart mobility [1] is one of the significant properties of smart cities which improvise the quality of life. Intelligent Transportation Systems (ITS) paved the way to realize smart mobility with the focus of efficient as well as safe navigation. However, the traffic accident statistics [2] insist that novel unconventional methods are inevitable in transportation engineering. Hence the modern ITS witnessed a strong rise in connected cars that are communicating with each

S.M. Thampi et al. (eds.), *Advances in Signal Processing and Intelligent Recognition Systems*, Advances in Intelligent Systems and Computing 678,
DOI 10.1007/978-3-319-67934-1_19

other using Vehicular Ad-Hoc Network (VANET). Here, the cars can be connected to any other vehicle, Road Side Unit (RSU), public infrastructure, human and sensors to get a spectacular overview of current situation than the conventional driver's purview.

The major connected car safety applications such as, collision warning systems [3–6] set aside vehicles to constantly monitor the state of nearby vehicles as well as road conditions to avoid accidents by taking appropriate timely decisions. Alternatively, extracting the Situation Awareness (SAW) from the sensor data and alert the drivers about the current traffic and road situation is still lacking in the highway context. Generally, SAW can be expressed as "the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning and the projection of their status in the near future" [7]. However, Situation Assessment (SA) represents the processes involved in information processing and their interaction with the other entities or environments to obtain their SAW. In concise, the SAW can be viewed as 'knowledge' whereas SA as 'processes' involved in obtaining that knowledge. Conversely, the most essential step in connected cars scenario is to ascertaining the association between SAW and driver's driving behaviors which is not fully explored and requires further investigation.

In this research work, an Adaptive Neuro-Fuzzy Inference System (ANFIS) [8] based Collision Warning System (CWS) is proposed for connected cars safety application. Firstly, the proposed model gathers the data from sensors which are implanted in the vehicles. For that, this research work uses the following sensors: ultrasonic sensor, gas sensor and accelerometer sensor. The ultrasonic sensor is used to identify the presence of any other vehicles in the front or rear side, gas sensor is used to recognize the alcohol consumption of driver and accelerometer is used to capture the wobbling of vehicles. Further, the proposed model assumes each vehicle has a Raspberry Pi board over which these sensors are integrated. Secondly, each vehicle (i.e. Raspberry Pi) transmits or receives the sensor data with other using either WiFi or VLC. Thirdly, the collected sensor data will be given as input to the proposed ANFIS and Mamdani fuzzy models which has fuzzy stability model and assesses the current stability level of the target car driver. Since the proposed CWS is an active safety application, it requires the active participation of the driver to manually execute the remedial action in response to alerts. The major contributions of the paper are:

1. ANFIS based CWS which assesses the stability of the car driver using the information obtained from sensors.
2. VANET platform that enables the communication among vehicles using either VLC or WiFi. Here, the VLC is used for communication (i.e. for speed and bandwidth) whenever the Line of Sight (LoS) is available between the source and sink vehicles and can dynamically switch to WiFi, if LoS is not available.

The rest of the paper is organized as follows: Sect. 2 highlights the important closely associated research works whereas the proposed architecture is explained in Sect. 3. The experimental set-up and the information flow is explained in Sect. 4. The important findings are consolidated in Sect. 5. Finally, Sect. 6 elucidates the conclusions as well as future directions of this research work.

2 Related Works

In this section, the research works that have close connection with the objective of the paper are highlighted. The three level connected cars scenario is well explained by L.Nanjie in [9] where the client layer contains the sensors that are collecting the data, connection layer illustrate the communication among cars and the cloud layer is used for processing the sensor data. According to M.Gerla [10], the vehicles are considered as mobile sensor platform which collects the sensor data and broadcast the appropriate data to the interested parties is explained by N. Lu [11]. Also, P.M. Salmon [12] highlighted that the methodology to associate the relationship between SAW and driver's carelessness is not yet matured and requires extensive investigation. M. Liggins [13] explains the significance of situation assessment in SAW. A. Eskandarian [14] proposed an active Advanced Driver Assistance System (ADAS) that avoids the traffic incidents by automatically acquiring the control of the car. Matheus et al. [15] recommend the semantic-based Situation AWareness Assistant (SAWA) to develop domain-specific ontologies for sensor data collection process. S. Sivaraman et al. [16] proposes an integrated ADAS to establish a association among driver, vehicle and infrastructure to have a complete view of hazardous situations. R. Hoeger et al. [17] proposed an ADAS architecture for automated vehicles in HAVEit project. R. Isermann et al. [18] proposed PRORETA project which is a collision avoidance system during traffic jam and sudden appearance of stationary objects which are common causes of accident.

3 ANFIS Based Situation Awareness Assessment for VLC Enabled Connected Cars

The major objective of the proposed work is to develop a CWS for connected cars. The proposed CWS continuously monitors the driving behaviors of proceeding or succeeding vehicle drivers using the sensor data obtained from those target vehicles (refer to Fig. 1). Later, these sensor data are analyzed by ANFIS to identify the harmful driving situations by assessing the stability level of drivers. Subsequently, it sends the alert messages to other vehicles. In concise, the proposed CWS contains computation as well as communication components. Here, the computation component does the ANFIS execution which installed and implemented over Raspberry Pi boards. However, the communication between the Raspberry Pi is accomplished using either WiFi or VLC.

3.1 Raspberry Pi

Being a Single Board Computer (SBC), Raspberry Pi supports bus interfaces for communicating with its external interfaces. Raspberry Pi is installed with 'scikit-fuzzy' and 'anfis' libraries. 'The scikit-fuzzy' is a robust fuzzy logic toolkit for SciPy that implements several fuzzy logic algorithms. Additionally, the 'anfis' is used for Adaptive Neuro Fuzzy Inference System.

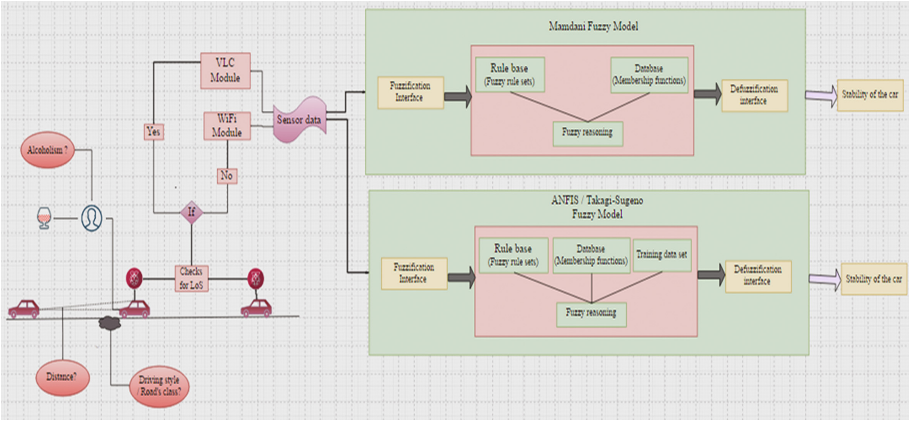

Alcoholism ?
VLC Module
WiFi Module
Yes
No
If
Checks for LoS
Sensor data
Distance?
Driving style / Road's class?
Mamdani Fuzzy Model
Fuzzification Interface
Rule base (Fuzzy rule sets)
Database (Membership functions)
Fuzzy reasoning
Defuzzification interface
Stability of the car
ANFIS / Takagi-Sugeno Fuzzy Model
Fuzzification Interface
Rule base (Fuzzy rule sets)
Database (Membership functions)
Training data set
Fuzzy reasoning
Defuzzification interface
Stability of the car

Fig. 1. Proposed Fuzzy Inference System

3.2 Sensors

The presence of sensors is inevitable in the context of connected cars. Sensors are used to collect the environment information whereas the Raspberry Pi is used to process these data to infer the stability of the driver. This research work uses information such as inter-vehicle distance, the angle of inclination of a vehicle with respect to the ground surface and alcohol trace in the breath to draw the conclusion on the stability level of a target vehicle driver. The inter-vehicle distance among its peer vehicles is measured using ultrasonic sensor (HC-SR04), the angle of inclination is calibrated using accelerometer sensor (MPU 6050) and the gas sensor (MQ5) is used to check whether the driver consumes alcohol or not.

Accelerometer Sensor - MPU 6050. MPU 6050 is a 6-axis integrated motion tracking device. Among those 6-axis, each accelerometer and gyroscope shares three axes. The accelerometer measures the linear acceleration of movement both due to the movement as well as due to gravity. Since the accelerometer readings are not that much sensitive when the tilt angle is greater than ±45°, all the three axes of the accelerometer are taken into account, which compensates one axes accuracy lag by another axis's accurate sensitivity. In addition to this, it also has gyroscope measures the angular rotational velocity. MPU-6050 supports I2C protocol so that the data can be easily communicated with the Raspberry Pi.

Gas sensor – MQ5. MQ5 is a gas sensor which even has a sensitivity to alcohol traces. It produces analog values and hence there is a need for analog to digital conversion (ADC) to get the digital values from the sensors. Since Raspberry Pi is fully digitalized in pins, an external ADC circuit is needed.

Ultrasonic sensor – HC-SR04. The HC-SR04 has an in-built transmitter, receiver and a controlling unit. The transmitter has a capability to transmit eight sonic beams from it, which is of range 40 kHz and the receiver is tuned to receive the reflected signal, which is due to the obstacle, of those signals. Before starting the eight sonic beam transmissions, the timestamp is taken and on the reception of the reflected signal, the timestamp is taken along. Based on the time taken between the transmission of the signal and the reception of the reflected signal, that is found using timestamps, the distance between the range finder and the obstructing vehicle can be found. These sensor data are then transmitted using VLC from one vehicle (V1) to its successor vehicle (V2), so that it can calculate the stability factor of V1 in V2.

3.3 Visible Light Communication

Transmitter. The LED is the most efficient source of light, which is a result of solid state lighting technology revolution. The cars usually have head and tail lights, which are usually made upon LEDs. As a result, the transmitter needed for VLC communication is established partially. Additionally, most renowned car manufacturers like Audi and BMW, have also implemented LASERs in their headlights. Consequently, the inter-vehicle communication distance can be extended beyond the known limits. The VLC transmitter consists of one Hand Shaking (HS) unit and one data Communication (Com) unit. The HS unit is used to ensure that two vehicles in a communication have LoS whereas Com unit does the actual data transmission. If HS unit does not return ‘1’ means those two vehicles does not have LoS and immediately instructs the Com unit to stop the transmission or vice versa.

Receiver. The VLC receiver detects the presence of the light and converts it to a voltage to decide whether the data is ‘1’ or ‘0’. In general, the VLC uses photo-detector as their receiver. However, this research work uses solar panels as its VLC receiver. Since the solar panels provide larger area, it is easy to maintain LoS compared to photo-detector in the context of connected cars. Solar panels works on the basics of photo voltaic cell theory which converts the light energy either from the sun or the other sources of light into electrical energy.

VLC in Action. The actual VLC communication among vehicles happens as follows (refer to Fig. 2): Here, each car is depicted as individual processing units, which is Raspberry Pi in this scenario. Car ‘A’ will collect all the above mentioned sensor data and transmit it to its predecessor car ‘B’ using VLC.

Each car will have a data communication unit and a handshaking unit in the front as well as rear side which makes the communication bidirectional in all the conditions. For communicating with the preceding vehicle it will make use of front-side communication unit and handshaking units. In contrast, to communicate with the succeeding vehicles, it will make use of rear-side communication unit and handshaking units. Here the communication unit can be head or tail light whereas the handshaking

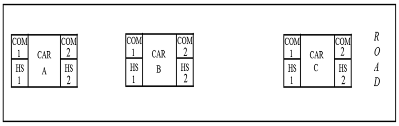


Fig. 2. VLC communication between vehicles

unit is the solar panels which can act as both energy harvester as well as VLC receiver. The on-off keying technique is use here for visible light communication. The data is simply transmitted by powering on and off the LED, which is the light source and transmitter for VLC. The powering on and off can also be altered by modifying the power supply so that the LED is not turned off completely rather its intensity is reduced. It is known as dimming support, which is offered by the LED. It is a simple and cost effective setup, which is easy to implement also. The sensor data which are transmitted from its predecessor vehicles are received by the successor vehicle via VLC and are processed by the techniques which are explained in the next section.

4 Fuzzy Inference System Implementation for Assessing Driver Stability

The system that makes use of fuzzy logic to compute the output from the obtained features is known as Fuzzy Inference system (FIS). The proposed SA for assessing the stability of the driver is implemented using both Mamdani and ANFIS models.

4.1 Mamdani Model

In this model, the crisp input from the sensor is fed into fuzzification which returns the fuzzy sets that represent the truthiness of the state, which is known as membership values. Here, the membership values are based on the membership functions, which tell the relationship between the values of the element and its degree of membership in that set. Based on the rule sets that are defined for a chosen membership function, these fuzzy sets are calculated. As per the rule strength, the fuzzy sets of all the output curves are added so that the resulted fuzzy set is calculated. After that, the de-fuzzification is applied which converts the fuzzy value to crisp value.

The proposed research work uses triangular (trimf) and trapezoidal membership function (trapmf). Additionally, the special forms of the triangular membership function are also used in this implementation. The membership function can be in any form of shape, which mostly based on the state of the input. This research work chooses the following membership function for each input variables as given below (refer to Fig. 3):

The Mamdani fuzzy model is a five step process:

Step 1: The feature crisp values are obtained from the sensors (MQ5, HC-SR04, MPU 6050) that states the nature of the predecessor vehicle. Those values are given as inputs to the step 2.

Step 2: The fuzzification is done to calculate the membership values using the membership functions. For instance, consider the tilt angle of the vehicle with respect to the ground surface can be in the range of 0° to ±90°.The tilt angle will be nearly equal to 0° and when it is inclined to the left side, it will give readings up to 90° correspondingly; when it is inclined to the right side it gives readings up to −90°.Hence the tilt angles are grouped into three groups namely smaller, medium and higher angles whose values are as −90° to 0°, −35° to 35° and 0° to 90° (refer to Fig. 3(a)). Similarly the

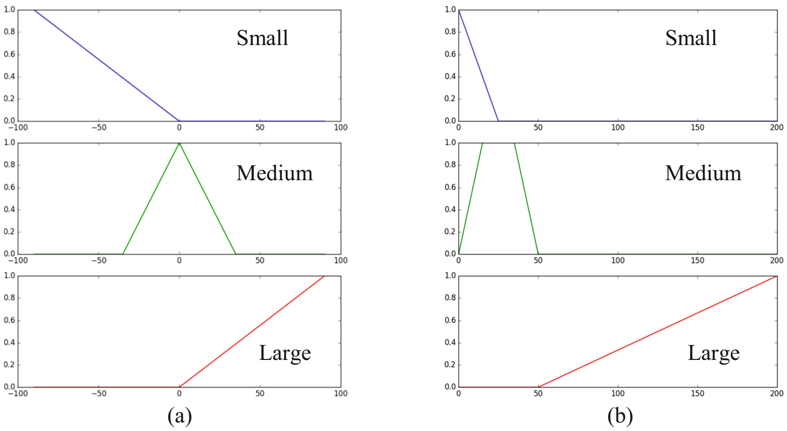


Fig. 3. (a). Membership functions for Tilt angle(Accelerometer). (b). Membership functions for inter-vehicle distance (Ultrasonic)

inter-vehicle distance uses both the triangular as well as trapezoidal membership function since, the trapezoidal function states the medium inter-vehicle distance for which the output should not be changed. For small and larger inter-vehicle distances the same triangular membership function is used. After defining the membership function for all the input variables, define the membership function for output variable also. Now, the interp_membership function is used to define degree of membership value and for −10.3°, it is found as 0.114444444444 (for low tilt angle mf), 0.705714285714 (for medium tilt angle mf) and 0.0 (for larger tilt angles). From these values, it states that the tilt angle −10.3° is having a higher probability of being a medium tilt angle rather than smaller tilt angle, while can never be larger tilt angle.

Step 3: After that, the set of rules are written to constrain the possible output levels as stated below:

Rule 1. When alcohol is low AND relative distance is high AND tilt angle is medium, Stability is high

Rule 2. When alcohol is low AND distance is medium AND tilt angle is medium, Stability is medium

Rule 3. When alcohol is high AND distance is high AND tilt angle is medium OR When alcohol is high AND distance is medium AND tilt angle is low OR tilt angle is high, Stability is low

Based on the input crisp value and degree of the fuzzy membership function, the respective rule sets are applied. Subsequently, resultant rule strength is mapped to output fuzzy set function to create the consequence for that rule.

Step 4: Thus obtained consequence sets are combined into single resultant output function that can be stated as below(refer to Fig. 4):

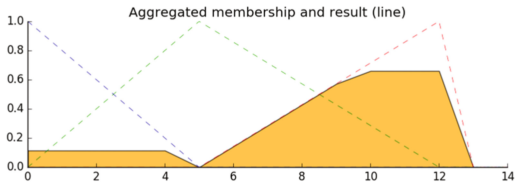


Fig. 4. Resultant fuzzy set after applying the defined rule sets

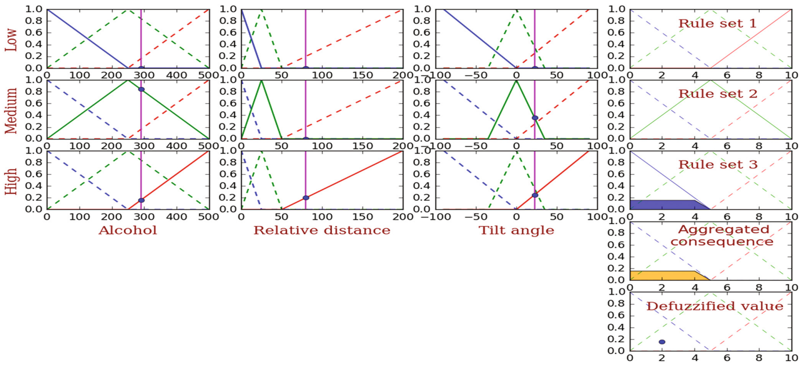


Fig. 5. Mamdani FIS process for Alcohol = 290, relative or inter-vehicle distance = 80 and Tilt angle = 22.3

Step 5: Finally, apply the defuzz function to convert this fuzzy set into crisp value which is then used in decision making. For that, there are several methods available for de-fuzzification: Centroid, maxima, weighted average, middle of maxima. This research work uses centroid as its defuzzification function to obtain the stability level of driver (refer to Fig. 5).

4.2 ANFIS Model

The major challenge of Mamdani model is the number of rules that are stated to make a decision about an incident is limited. Hence, the best practice is to use all possible combinations of the rules that can be obtained. This approach helps to narrow down the missed-out case situations. Further, the system can be made adaptive in a way the erroneous terms can be modified using suitable optimization algorithms. ANFIS is one such model whose architecture is given below. Basically, the ANFIS is a machine learning approach with five intermediate layers which obviously contains two phases: training and testing phase.

Training phase. The ANFIS training architecture is given in Fig. 6 and are explained as below.

Layer 1. In this layer, the training datasets are given as inputs to the membership functions to denote the levels of the premise parameters that are responsible for those membership functions generation. For example, for a given Gaussian membership function, the mean (μ) and sigma (σ) are premises parameters on which the membership functions shape truly rely on. Based on these premise parameters and the given training input dataset, the degree of truthiness for all the inputs corresponding to all the desired levels of the inputs are found and denoted as,

$$O_1 = \mu_i(x) = e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \qquad i = A,\ B,\ C$$

Where $\mu_i(X)$ represents the degree of truthiness for one of the given inputs like alcohol (A), relative distance (B) and tilt angle from accelerometer (C). The number of nodes in this layer will be equal to the product of the number of each individual inputs membership levels considered in implementation.

Layer 2. This layer decides the firing strength of desired rules. To rule out the missing conditions, the combinations for all the levels of the given inputs are considered. This is labeled as *π*, as it represents the product of the membership values for all the inputs. The output of each node in this layer can be mathematically represented as follows. The output of these nodes is also known as firing strength of desired rules.

$$O_{2,j} = w_j = \mu_{Aj}(x) * \mu_{Bi}(y) * \mu_{Ci}(z), i = 1,\ 2,\ 3\,\text{and}\,j = 1,\ 2\ldots27$$

Layer 3. This layer calculates the normalized firing strength of each node represented in layer 2. This can be mathematically represented as

$$O_{3,j} = \overline{w_j} = \frac{w_j}{\Sigma w},\ \text{where}\,j = 1, 2.\ldots27$$

Layer 4. The layer 4 implements the real adaptive nature of the Adaptive Neuro-Fuzzy Inference Systems. The nodes output is represented by an adaptive mathematical function which can be represented as

$$O_{4,j} = \overline{w_j}.f = \overline{w_j}\left(p_j x + q_j y + r z_j + o_j\right),\ \text{where}\,j = 1, 2.\ldots\ldots27$$

The parameters (x, y, z) represents the input sets of the three distinct inputs and ‘o’ represents the desired output that has to be obtained for this combination of the inputs. The p, q, r represents the weightings for the input sets, so that the desired output can be achieved with or without minimum deviation from that of the desired output. These parameters are known as Consequent parameters.

Least Square Error. The values of consequent parameters are obtained and optimized with the help of least mean square optimization algorithm before continuing with layer 5. Simply it will consider the layer 4 node's output and desired output and from that, the deviation is calculated by least square error algorithm.

$$[\mathrm{O}_4][\mathrm{X}] = [\mathrm{O}]$$

Where, $[\mathrm{O}_4]$ represents the matrix formatted layer 4 output, [X] represents efficient coefficient for this representation, O represents the matrix formatted overall desired output. The co-efficient can be found using simple inverse matrix whose solution can be represented as,

$$[\mathrm{X}] = [\mathrm{O}_4]^{+}[\mathrm{O}], \text{ where } [\mathrm{O}_4]^{+} = \left([\mathrm{O}_4]^{\mathrm{T}} \cdot [\mathrm{O}_4]\right)^{-1} . \mathrm{O}_4^{\mathrm{T}}$$

This is a simple notation for obtaining the co-efficient but the operation of finding the inverse of a matrix is very complex in nature and it will consume huge computation time. Hence, the simplified notation for obtaining the same result with lesser computation cost is represented as follows.

$$[\mathrm{X}_{\mathrm{i}+1}] = [\mathrm{X}_{\mathrm{i}}] + \mathrm{S}_{\mathrm{i}+1}.[\mathrm{O}_4]_{\mathrm{i}+1} \cdot \left([\mathrm{O}]^{\mathrm{T}} - [\mathrm{O}_4]^{\mathrm{T}}.[\mathrm{X}_{\mathrm{i}}]\right), \text{ where i = No of training data set}$$

$$[\mathrm{S}_{\mathrm{i}+1}] = [\mathrm{S}_{\mathrm{i}}] - \frac{[\mathrm{S}_{\mathrm{i}}].[\mathrm{O}_4]_{\mathrm{i}+1}.[\mathrm{O}_4]_{\mathrm{i}+1}{}^{\mathrm{T}}.[\mathrm{S}_{\mathrm{i}}]}{1 + [\mathrm{O}_4]_{\mathrm{i}+1}{}^{\mathrm{T}}.[\mathrm{S}_{\mathrm{i}}].[\mathrm{O}_4]_{\mathrm{i}+1}}$$

Where the matrix S is known as co-variance matrix and the initial conditions before performing the training, it is set as $[\mathrm{S}_{\mathrm{o}}] = \gamma \mathrm{I}_{\mathrm{p}}$ where γ is the positive large number and I_{p} represents the identity matrix of the dimension of the product of a number of inputs plus one and number of rule sets obtained. The value of [X] is refined for each number of training data set and hence the efficient co-efficient can be obtained.

Layer 5. Finally, this layer predicts the error rate in achieving the desired output by comparing the obtained output with the desired one. The obtained result is based on the value obtained from LSE prior to layer 5. Based on the obtained error value, if it is less than the prescribed threshold error or lesser than the defined epochs, the premises parameters are modified to obtain better adaptation. This adaptation is based on the gradient descent approach, which is a strong optimization technique.

Back propagation method: The one way to reduce the error rate is to minimize the mean of that parameter itself. The updated formulae for generic premise parameter are given below, where ♒ is the learning rate and k is the step size.

$$\Delta\alpha = -\eta \frac{\partial E}{\partial \alpha}, \text{ where } \eta = \frac{\mathrm{k}}{\sqrt{\sum_{\alpha} \left(\frac{\partial \mathrm{E}}{\partial \alpha}\right)^2}}$$

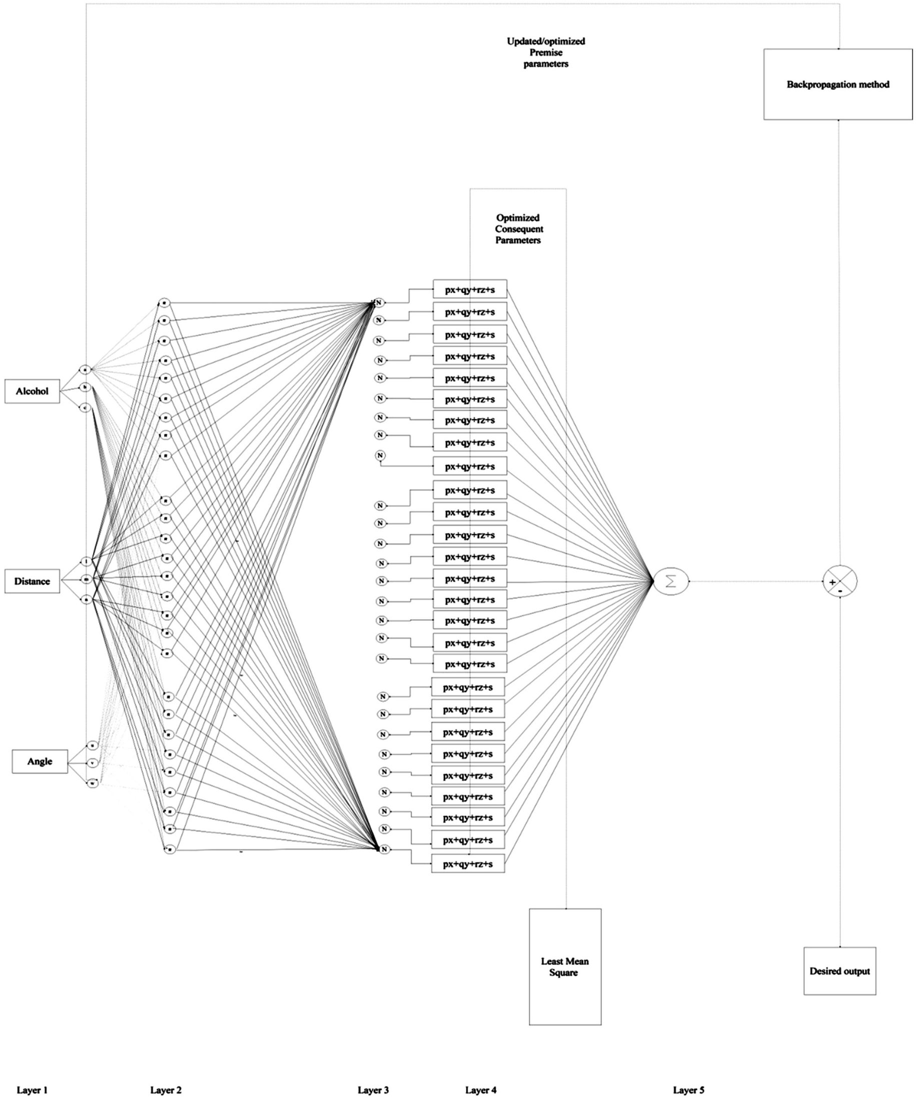


Fig. 6. ANFIS training architecture

Prediction phase. The architecture is given in Fig. 7. When the model is successfully trained as stated above, it is ready to predict the nature of the stability by using the training that it has undergone. During the prediction state, the adaptive nature or optimizations phases of ANFIS will never be applied. Basically, all these actions detailed in layer one to five, LSE and back-propagation methods will be applied during training phase whereas the actions detailed in layer one to five will be implemented in stability prediction phase. As the training is the key role in any kind of machine learning, so does here also. The nature of training always has an impact on the quality

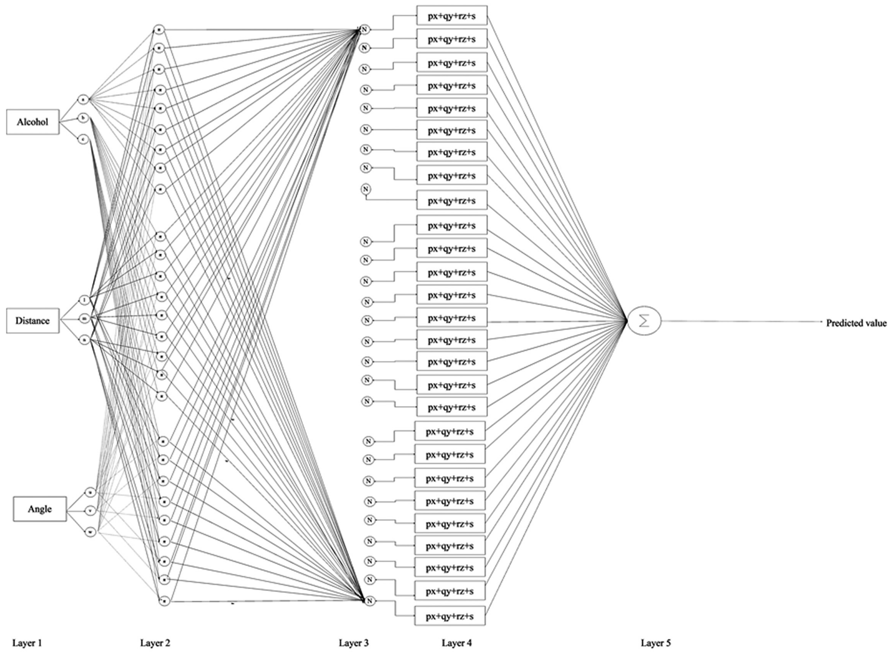


Fig. 7. ANFIS predicting architecture

of prediction as well as time taken for training. The time taken to train the ANFIS model over Raspberry Pi, Laptop and the comparison between Mamdani and ANFIS model are given next section.

5 Results and Discussion

This research work proposes a CWS for connected cars safety application. The proposed CWS is validated using both Mamdani and ANFIS model. This section is going to compare both of them and highlight the merits and demerits of each model. Firstly, the fuzzy logic is more dependent on the user specified rule sets, the crisp value that was given by the fuzzy may not be up to the mark for some cases, as the developer may have missed out some chances to define those areas. As a net result, the effectiveness is brittle in Mamdani fuzzy systems. On the other hand, the time needed for stating the stability is very small (in very few seconds). Additionally, the Mamdani fuzzy logic does not require training. In contrast, the ANFIS takes every single combination into consideration and hence, there is no chance for missing of rules and the approximate result can be always obtained at anytime. However, the cost of training is linearly related to its training dataset size. ANFIS consumes large amount of time for training whereas it took only few seconds of time to predict the stability.

Table 1 compares the stability values of both Mamdani and ANFIS models for the same sensor input data. Here, lesser the SAW value, driver has more stability.

Table 1. Comparison of Mamdani and ANFIS model for driver stability

Inputs from environment			SAW output values	
Alcohol	Relative distance	Tilt angle	Fuzzy logic	ANFIS
9	1	0.4	4.542	2.92948262
197	110	20	7.7629157	6.69371058
447	110	15	2.2592592	4.80595243
194	190	10	7.7831978	8.95677947
197	109	19	7.7629157	6.75022173
229	12	2	2.2592592	3.60587773
300	12	2	2.2592592	3.59996461
148	99	09	7.9048800	7.55720714
459	120	19	2.2386587	4.66735067
290	167	23	2.259259	6.67582136
470	23	09	2.259259	2.808469
398	120	9	2.259259	5.68921093
297	98	36	2.259259	4.25744907

From Table 1, it is evident that the SAW value obtained from both the prediction algorithms are close enough to each other. However, the Mamdani model is not that much responsive for minor changes in the environments, which makes vulnerable the prediction in real world scenarios when compared with high responsive ANFIS prediction model.

6 Conclusion and Future Work

Situation assessment is an important ingredient of intelligent transportation system that constantly monitors the state of nearby vehicles as well as road conditions to avoid accidents by taking appropriate timely decisions.

With situation awareness and situation assessment, the driver may get a clear view about the present driving situations and be able to adopt timely decisions to circumvent the forthcoming dangers. This research work proposes a collision warning system which analyzes the stability of target vehicle driver using Mamdani and ANFIS fuzzy models. Though the Mamdani model has a quick response time while computing the stability of the driver, it has less sensitivity and precision for a given sensor data. In contrast, the ANFIS is a robust model that has high sensitivity, high precision and low response time. But it needs training for their membership functions that incurs additional overhead in the model. Further, the inter-vehicle communication in the proposed model is realized using visible light communication thereby exploiting the speed and bandwidth of VLC during data transmission. To enable reliable inter-vehicle communication, the proposed research work uses WiFi for data communication, if the LoS is not maintained between source and sink vehicles. In future, the proposed active CWS can be modified to passive CWS by integrating deep learning, information fusion and summarization thereby automating the decision making and controlling of vehicles.

References

1. Djahel, S., Doolan, R., Muntean, G.-M., Murphy, J.: A communications-oriented perspective on traffic management systems for smartcities: challenges and innovative approaches. IEEE Commun. Surv. Tutor. **17**(1), 125–151 (2015)
2. http://www.asirt.org/. Accessed May 2015
3. Girard, A.R., de Sousa, J.B., Misener, J.A., Hedrick, J.K.: A control architecture for integrated cooperative cruise control and collision warning systems. In: Proceedings of the 40th IEEE Conference on Decision and Control, vol. 2, pp. 1491–1496. IEEE (2001)
4. Meguro, J.-I., Murata, T., Takiguchi, J.-I., Amano, Y., Hashizume, T.: GPS multipath mitigation for urban area using omnidirectional infrared camera. IEEE Trans. Intell. Transp. Syst. **10**(1), 22–30 (2009)
5. Huang, J., Tan, H.-S.: Error analysis and performance evaluation of a future-trajectory based cooperative collision warning system. IEEE Trans. Intell. Transp. Syst. **10**(1), 175–180 (2009)
6. Polychronopoulos, A., Tsogas, M., Amditis, A.J., Andreone, L.: Sensor fusion for predicting vehicles' path for collision avoidance systems. IEEE Trans. Intell. Transp. Syst. **8**(3), 549–562 (2007)
7. Endsley, M.R.: Design and evaluation for situation awareness enhancement. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting, vol. 32, pp. 97–101. SAGE Publications (1988)
8. Jang, J.S.R.: ANFIS: adaptive-network-based fuzzy inference system. IEEE Trans. Syst. Man Cybern. **23**, 665–685 (1993)
9. Nanjie, L.: Internet of vehicles your next connection. Technical report (2011)
10. Gerla, M., Lee, E.K., Pau, G., Lee,U.: Internet of vehicles: from intelligent grid to autonomous cars and vehicular clouds. In: IEEE World Forum on Internet of Things. IEEE (2014)
11. Lu, N., Cheng, N., Zhang, N., Shen, X., Mark, J.: Connected vehicles: solutions and challenges. IEEE Internet Things J. **1**(4), 289–299 (2014)
12. Salmon, P.M., Stanton, N.A., Young, K.L.: Situation awareness on the road: review, theoretical and methodological issues, and future directions. Theor. IssuesErgon. Sci. **13**(4), 472–492 (2011)
13. Liggins, M., Hall, D., JamesLlinas, P.: Handbook of multi-sensor data fusion: theory and practice. The electrical engineering and applied signal processing series. CRC Press Inc., Boca Raton (2009)
14. Eskandarian, A.: Fundamentals of driver assistance. In: Eskandarian, A. (ed.) Handbook of intelligent vehicles, pp. 491–535. Springer, London (2012)
15. Matheus, C.J., Kokar, M.M., Baclawski, K., Letkowski, J.A., Call, C., Hinman, M.L., Salerno, J.J., Boulware, D.M.: SAWA: an assistant for higher-level fusion and situation awareness. In: Dasarathy, B.V. (ed.) Multisensor, Multisource Information Fusion: Architectures, Algorithms, and Applications 2005. Society of PhotoOptical Instrumentation Engineers (SPIE) Conference Series, vol. 5813, pp. 75–85 (2005)

16. Sivaraman, S., Trivedi, M.: Towards cooperative, predictive driver assistance. In: Proceedings of the16th International IEEE Conference on Intelligent Transportation Systems, ITSC, pp. 1719–1724 (2013)
17. Hoeger, R., Amditis, A., Kunert, M., Hoess, A., Flemish, F., Krueger, H.P., Bartels, A., Beutner, A., Pagle, K.: Highly automated vehicles for intelligent transport: have-it approach. In: Proceedings of the 15th World Congress on Intelligent Transport Systems and ITS America's 2008 Annual Meeting (2008)
18. Isermann, R., Mannale, R., Schmitt, K.: Collision-avoidance systems PRORETA: situation analysis and intervention control. Control Eng. Pract. **20**(11), 1236–1246 (2012). Special Section: Wiener-Hammerstein System Identification Benchmark

Speech Recognition Using Feed Forward Neural Network and Principle Component Analysis

Nusrat Momo, Abdullah, and Jia Uddin(✉)

BRAC University, Dhaka 1212, Bangladesh
Nusratsuzanamomo@gmail.com,
abdullah.5121011@gmail.com, engrjiauddin@gmail.com

Abstract. Various models have been proposed with many dimension reduction techniques and classifiers in the field of pattern recognition by using audio signal processing. In this paper, an effective model has been proposed for pattern recognition using PCA as the sole dimension reduction technique and Feed forward Neural network as the classifier. Twenty-eight Parkinson's disease affected patients' audio recordings consisting of the pronunciation of the vowels 'A' and 'O' have been used as the dataset. From these audio recordings twenty features were extracted and PCA was run on those features. PCA rearranged the feature vector matrix in a more optimized manner. Thus the optimal features were arranged in order of their significance. From this rearranged and optimized feature vector matrix, the first eight optimal features were chosen which were later used to train and test the classifier Feed forward Neural network. Experimental results demonstrate that the model can predict the occurrence and pattern of the vowels 'A' and 'O' from the audio files with very high accuracy compared to the swarm search for feature selection in classification.

Keywords: Feature extraction · Speech recognition · Feed forward neural network

1 Introduction

The Parkinson's disease (PD) is a fatal disease that has a long-term effect on a person's life. It is a chronic disease that attacks the motor system of the central nervous system of a person. PD causes drastic changes for instance—a PD affected person experiences shaking, difficulty in walking, depression, behavioral problems and more significantly it can influence a person's voice, making them talk very slowly or experience difficulties pronouncing words. A normal person cannot relate to a PD affected patient's struggle. Thus, their speech becomes less understandable for other healthy persons who are not suffering from PD. As a result, it becomes difficult to distinguish the letters especially the vowels in their speech. So, the main agenda of this proposed model is to properly recognize the pattern, more specifically the vowels 'A' and 'O' in their speech. The first and foremost step to achieve that is to extract the features of the corresponding audio files as the features will help to distinguish the vowels that are being pronounced. But only extracting the features from their speech will not give an optimal result with

S.M. Thampi et al. (eds.), *Advances in Signal Processing and Intelligent Recognition Systems*, Advances in Intelligent Systems and Computing 678,
DOI 10.1007/978-3-319-67934-1_20

high accuracy. Therefore, the selection of the features which are more comparable and optimal is necessary to help differentiate the vowels ‘A’ and ‘O’ with greater precision. Moreover, this feature selection will help the classifier to recognize the vowels easily. For pattern recognition, feature extraction is a key necessity. In this area, a classic and quite popular feature extraction method, Principal component analysis (PCA), also known as Karhunen-Loeve expansion is used quite frequently [9]. There are some significant advantages of PCA. For instance, because it takes mutually orthogonal axis there is less redundancy of information without any loss of information in the original dataset [2]. Moreover, it maximizes variance thus it reduces noise [14]. Computations for large datasets can be done easily with PCA. Furthermore, PCA produces results quite fast in a very efficient manner.

Coming to the classifier, Neural network is extensively used in pattern recognition as a classifier. It requires minimal statistical training and the whole process is relatively simple and easy to use. Complex nonlinear relationships between dependent and independent variables can also be detected by the neural network. Moreover, it can approximate any function irrespective of its linearity. Therefore, both Neural network and PCA had been used in the model to utilize their advantages.

The rest of the paper is organized as follows: Sect. 2 represents an overview of all the algorithms used in the proposed model, Sect 3 describes the implementation part, the methodology and work flow, Sect. 4 shows the result analysis and corresponding histograms and error graphs and finally, Sect. 5 concludes this paper.

2 Algorithm Overview

In this section, the two main algorithms that have been used for the model will be discussed. Principle component analysis was used for dimension reduction and Feed forward neural network was used for pattern recognition.

2.1 Principle Component Analysis (PCA)

Principal Component Analysis (PCA) is a simple linear transformation algorithm and a statistical procedure which generally uses an orthogonal transformation. PCA searches for the maximum variance in the original space. Moreover, PCA is heavily used in feature selection method. The concept of eigenvector and eigenvalue is important in PCA analysis. PCA can be applied by the decomposition of eigenvalues of a data covariance matrix usually after mean centering and normalizing the data matrix for each attribute [1]. From a set of dataset eigenvectors and their corresponding eigenvalues can be determined as eigenvectors and eigenvalues exist in pairs. An eigenvector is basically a direction and its corresponding eigenvalue is a number or a value which indicates the variance of data in a direction. The eigenvector with the highest eigenvalue is considered as first principal component and the second highest eigenvalue with the next highest variation is the second principal component and following this pattern, it arranges the features [13].

2.2 Feed Forward Neural Network

Neural networks for continuous space language models have been attractive in recent years [19]. Neural networks are also known as connectionist systems. This technique is basically a computational approach utilized as a part of computer science. There are two important concepts of neural networks one is the perceptron and the other is sigmoid neuron. Roused by prior work by Warren Mcculloch and Walter Pitts perceptrons were developed by Frank Rosenblatt [11]. Perceptron is based on layers and does the computation by using easy addition and subtraction. Frank Rosenblatt explained about circuitry, for example, exclusive-or circuit with mathematical terms and notations but not in basic perceptron [15]. By using several binary inputs, a single perceptron gives one binary output. In the diagram, it is taking three inputs x1, x2, x3 and produces an output. Figure 1 shows a perceptron [6].

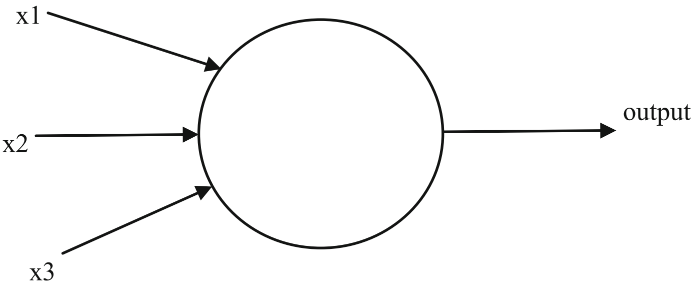


Fig. 1. Perceptron with 3 inputs x1, x2, x3 and 1 output

A perceptron works by computing outputs. For computing outputs, weights are used. The output either 0 or 1 is computed by a threshold value. If the weighted sum $\sum_j wjxj$ is greater than or less than a threshold value the output is either 0 or 1. The threshold value and the weight values are real numbers [17].

$$output = \begin{cases} 0 & if \sum_j wjxj \le thresshold \\ 1 & if \sum_j wjxj > thresshold \end{cases} \tag{1}$$

One problem with perceptron is a small change in the weights of any perceptron can affect the output of that perceptron to change completely from 0 to 1. This problem may have an adverse effect on the rest of the network and it may complicate the whole network. A sigmoid neuron can overcome this problem. They are a modified version of perceptron so a slight change in weights and bias does not have a major effect on the output. Like perceptron it will take x1, x2, x3 inputs but these inputs can take fractional values too, for instance, all values between 0 to 1. Moreover, the concept of weights and bias are also used in the sigmoid neuron. The output is σ (w·x + b), σ is known as the sigmoid function. The more explicit output can be defined by the following formula [12].

$$\frac{1}{1+e^{\left(-\sum_j wjxj\right)}} \tag{2}$$

Another type of neural network is feed forward neural network. This network is quite simple and easy as the data or information only move forward starting from the inputs units through the hidden units and lastly to the output units creating no cycles in the network. Feed forward neural networks predict answers quite well as they are universal approximators [18]. Continuous mapping can be done by a three sometimes a four-layered network using sigmoid saturating unit output functions even when there are many hidden parts [3, 5, 7]. Because of these advantages feed forward neural network has been used in the suggested model.

3 Methodology and Work Flow

Figure 2 illustrates a detail block diagram of our proposed model that consists of extracting features from 2D audio signal, select optimal features using PCA and train and test using Neural Network classifier.

The Algorithmic view of the proposed methodology is shown below.

```
proposedAlgorithm( List of Audio Files )
    featureMatrix = null;
    labelMatrix = null;
    i = 0;
    for each File
        r = readFile( File );
        featureMatrix.add( extractFeatures( r ) );
        labelMatrix[0][i] = 1; // 1 for 'a' and 0 for 'o'
        labelMatrix[1][i] = 0; // 1 for 'a' and 0 for 'o'
    end
    [coeff, score] = pca( featureMatrix );
    for n= 3:8
        pcaFeatureMatrix = score(:, 1:n)*coeff(:, 1:n)'
        create neuralNetwork( 30 ); // 30 hidden layers
        neuralNetwork( pcaFeatureMatrix', labelMatrix );
        plot( Performance graph );
        plot( Error histogram );
        plot( Training State graph );
        plot( Receiver Operating Characteristic );
        plot( Confusion matrix );
    end
end
```

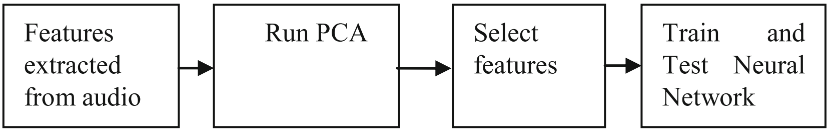


Fig. 2. Block diagram of proposed model.

3.1 Feature Extraction

Audio signals have many properties for example - amplitudes, maximum amplitudes, frequencies and length of sound and much more. Features are numeric properties of a signal. It is a measurable property that has been observed in phenomenon [8]. Different signals have different features, which may or may not be unique. For instance, there can be two separate signals with the same average amplitude. Therefore, it is very important to extract as many features as possible so that the most distinct features among them can be worked out. As these statistics play an important role in pattern recognition and machine learning, having as many as possible features increase the accuracy of predicting the correct letters. At the beginning, the research six features were extracted, namely Energy Entropy, Signal Energy, Zero Crossing Rate, Spectral Rolloff, Spectral Centroid, and Spectral Flux. These features together were not enough to accurately distinguish between the two vowels. The features are not distinct enough to recognize the correct vowel since there are many overlapping features in both the audios. Hence more features were needed to be extracted, which lead to the extraction of the following 14 features: root mean square (rms), standard deviation (std), kurtosis, maximum amplitude, minimum amplitude, mean amplitude, median amplitude, mean frequency, median frequency, skewness, peak to peak-the difference between maximum and minimum peak, peak to rms, root sum of square (rssq) and variance. These features along with the previous 6 features made the vowels more differentiable by the classifier. Obviously using all the features together do not give the highest accuracy because there are features that have values, which are common to both audios files.

3.2 Optimal Feature Selection

The feature vector is a matrix or vector that contains all the extracted features. If the number of features extracted from an audio file is n, each audio could be expressed as an n dimensional vector. If the audios consist the pronunciation of the same letter, features appearing in them will be similar. Pronounced features can be selected from the available feature to build an effective feature vector [10]. Again, to train the classifier, the most prominent features were needed to be selected. Therefore, the next step was to use dimension reduction algorithm, PCA. The features with the most differentiating values are acknowledged as feature vectors and stored in a matrix [13]. For the research, pca function was used. The function takes in a Matrix A of n × p and returns two matrixes namely coefficient matrix and score matrix. The coefficient matrix is a p × p matrix containing the principle component with the highest variance in the first column, the second highest variance in the second column and so on.

The score matrix is an n × p matrix which is a representation of Matrix A in the principal component space. The following code returns the coefficient and score matrix.

```
[coef, score] = pca(A)
```

To get our desired matrix with the features ordered by their uniqueness with the most distinguishable at the beginning column, the following code was used:

```
score(:, 1:n)*coef(:, 1:n)’
```

This returns the original matrix with the features ordered by their significance. Here n is 20 as there were 20 features. After applying this to the two individual matrixes where one contains all features of the audios of the letter ‘A’ and the other matrix containing the features of the audios of letter ‘O’, the two matrixes were combined into one matrix of 140 × 20. Let us call this matrix ‘T’ throughout the rest of the paper. The matrixes were combined in a random manner. For example, the first 5 rows are from the matrix of letter ‘A’ and then the next 8 are from letter ‘O’ and are continued with similar pattern. The next step was to take a different number of columns of the T matrix and find out the highest number of columns needed to achieve a high accuracy. Starting with the first 3 columns of T (140 × 3) and incremented till the desired accuracy was met. Table 1 shows a fraction of the feature values obtained after running PCA on the features mentioned in Sect. 3.1.

Table 1. Values of some features extracted.

F1	F2	…	F7	F8	……
0.075505085	−0.037215778	…	−0.201551897	0.090624379	……
0.060702219	−0.009612609	…	−0.208227243	−0.048785365	……
0.065863426	−0.02323271	…	−0.174564502	0.014228149	……
0.302426791	−0.198342311	…	−0.039702566	0.105251611	……
0.167757089	0.344376096	…	0.013973929	0.085328876	……

3.3 Training and Testing Using Neural Network Classifier

For classification, neural network classifier was used implementing the Feed forward algorithm. The classifier takes two matrixes, one that contains the feature vector and another that contains the labels corresponding to the row feature vector. The feature vector must have the features as rows and the samples as the columns. Therefore, the 140 × 3 matrix had to be transposed to achieve required 3 × 140 matrix. The label matrix has 2 rows and 140 columns. The first row contained the values 0 and 1 for letter ‘A’. One in the column that corresponded to the feature of ‘A’ in the feature vector and zero for the column that corresponded to the feature of ‘O’ in the feature vector. Similarly, the second column was for the letter ‘O’. After loading the two matrixes into the network containing 30 hidden layers. The network was trained with 30% of the data and the rest 70% were used for validation and testing. The process was repeated for the following matrixes T (4 × 140), T (5 × 140), T (6 × 140), T (7 × 140), and T (8 × 140). The network returned the following output for each matrix Performance graph, Error histogram, Training State graph, Receiver Operating Characteristic (ROC) and Confusion matrix. Using the above graphs, the minimum best number features to use from the matrix T to get the highest accuracy was determined.

4 Result Analysis

For this research, 140 de-noised audio recordings of 28 patients of Parkinson's disease are used [16]. The recordings are of letter 'A' and 'O'. From these audio files 20 features were extracted and using PCA, features were arranged by their significance. Using selective features from this matrix, the Neural Network was trained and tested. For determining the optimal number of features to get the highest accuracy, the following graphs Performance graph, Error histogram, Receiver Characteristic (ROC) and Confusion matrix for first 3 to 8 features, were obtained. The performance graph shows us training error, validation error and testing error in a graph and gives as an insight of any occurrence of overfitting. The error histogram reveals any outliers in the data set. Training state graph tells us about gradient value. The receiver operating characteristic checks the quality of the classifier. Confusion matrix sums up the accuracy and the performance to the whole system.

As for the performance, with each inclusion of a feature the validation curve and the testing curve gets closer to each other, implying that less of overfitting is done to the data. Not only overfitting decreases, also the cross-entropy decreases as well. But there is an exception with T (6 × 140) where there is a lot of significant overfitting and an increase in cross-entropy. For T (7 × 140), the overfitting decreased but the decrease in cross-entropy is not that significant. This has occurred due to outliers in data, as shown in the error histogram below.

The error histogram paints a clear picture of how the errors decrease, as more of data for outliers are fed to the classifier for training. As for T (6 × 140) and T (7 × 140) there are quite a few outliers at errors −0.02016, 0.02016 and at −0.05577, 0.05577 respectively. Nevertheless, the range of error has decreased from [−0.9365, 0.9365] in T (3 × 140) to [−0.00131, 0.00131] in T (8 × 140).

The graphs clearly show that the ROC curves moves closer to the y-axis, from T (3 × 140) to T (5 × 140) and stays vertical with the y-axis for the rest of the matrixes. This indicates that the quality of the classifier is very good.

The confusion matrix ensures that with an increase in a number of features the accuracy increases significantly. The model could achieve a very high accuracy with 5 to 8 features. The initial the overall accuracy with 3 features was 67.1% which increased drastically as more features were taken in consideration. The classifier was able to classify the two vowels properly. All the test and validation cases are being distinguished with their proper class.

As depicted in Fig. 3, each epoch the cross-entropy decreases. At 26 epoch the lowest cross-entropy, 1.8809e-05 is reached. There is no overfitting as the validation and test curves are close to each other which also implies that there are no outlier values in the data set. This is confirmed by the error histogram.

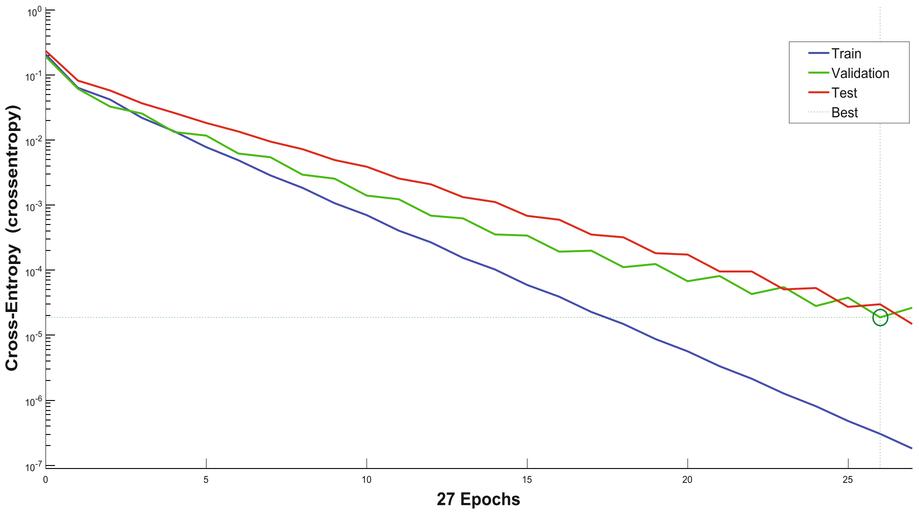


Fig. 3. Performance graph of T (8 × 140)

Figure 4 shows that all errors fall between −0.00257and 0.002575. The diagram clear shows that only a few data fall in −0.00149, −0.00041, 0.000407 and 0.001491. The rest majority are in the center. Hence there is less extrapolating of the data.

For checking the classifier's quality, the receiver operating characteristic is used widely. Threshold values are applied by roc over the interim [0, 1] for every classifier. The True Positive Ratio (TPR) and the False Positive Ratio (FPR) are the two values

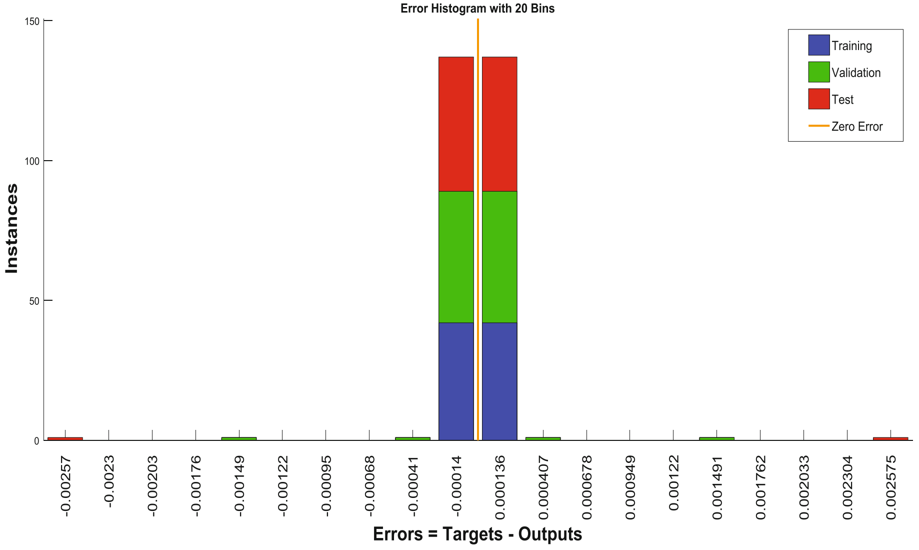


Fig. 4. Error histogram of T (8 × 140)

calculated for every threshold value applied by roc. For example, for a specific class, TPR stands for the number of outputs which are both actual and predicted class will be that specific class. However, for FPR the case is not the same. It stands for the number of outputs whose predicted class will be that specific class but the actual class will be different than that class. For TPR it must be divided by the number of outputs whose predicted class will be that particular class. But for FPR it must be divided by the number of outputs whose predicted class will not be that specific class. The area under the curve (AUC) gives a clear indication of the quality of the classifier. Figure 5 shows the curve is closer to the y-axis has an AUC of closer to 1. This indicates a better classifier. The more the curve is away from the y-axis, the worse is the classifier.

The matrix in Fig. 6 clearly tells us that how much of the predicted data matches the actual data. It gives the percentage of the predicted data are true positive, true negative, false positive and false negative. The green boxes are the place where the true positives and true negatives are, in other words, these are the percentages where the predicted ‘A’ matches with the actual ‘A’ and the predicted ‘O’ matches the actual ‘O’. The red boxes are the mismatch, were predicted ‘A’ is actual ‘O’ and vice versa. The blue box shows the overall accuracy and inaccuracy of the classifier. The system is highly accurate when using 8 features.

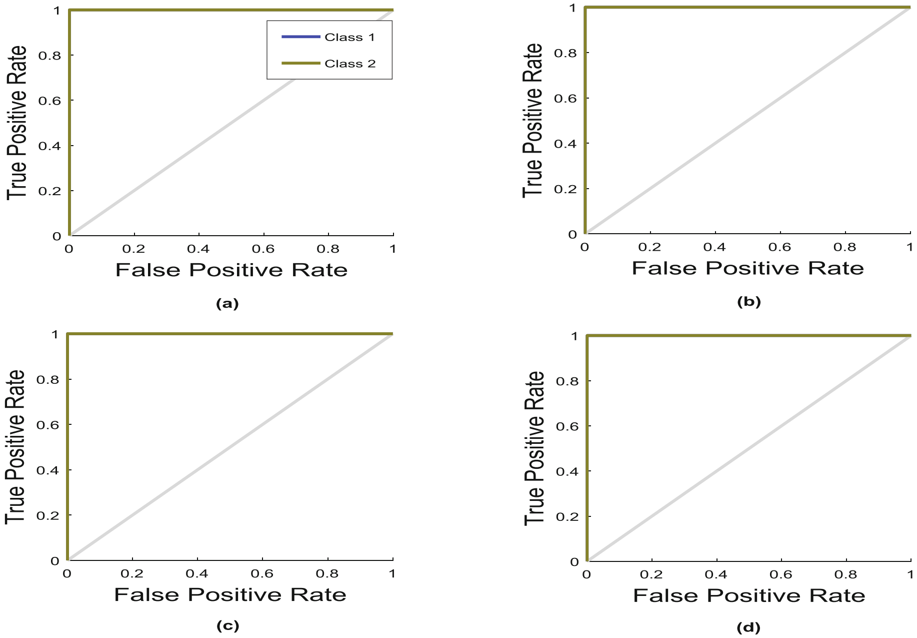
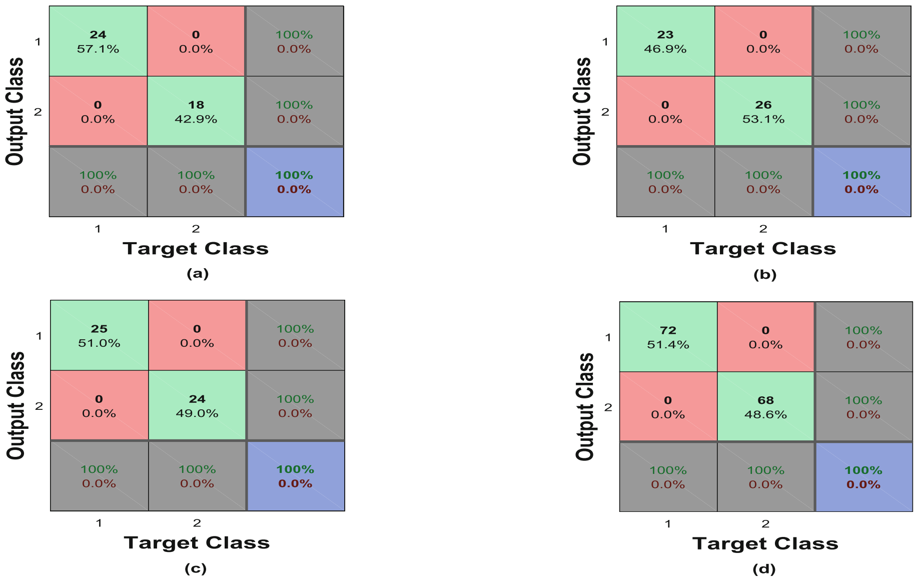


Fig. 5. (a) Training receiver operating characteristic of T (8 × 140), (b) Validation receiver operating characteristic of T (8 × 140), (c) Test receiver operating characteristic of T (8 × 140), (d) All receiver operating characteristic of T (8 × 140)



(a)

(b)

(c)

(d)

Fig. 6. (a) Training confusion matrix of T (8 × 140), (b) Validation confusion matrix of T (8 × 140), (c) Test confusion matrix of T (8 × 140), (d) All confusion matrix of T (8 × 140)

Table 2 shows a Comparison between our proposed model and other models [4]. As the Table 1 illustrates clearly that our model has a very low error rate compared to other models.

Table 2. Model comparison

Algorithms	Average error rate
Model 1	0.3126
Model 2	0.3250
Model 3	0.3210
Our Model	1.8809e-05

5 Conclusion

The model that has been proposed in this paper is quite effective with great accuracy. First, twenty features have been selected form the audio videos. They are Entropy, Signal Energy, Zero Crossing Rate, Spectral Rolloff, Spectral Centroid, and Spectral Flux. root mean square (rms), standard deviation (std), kurtosis, maximum amplitude, minimum amplitude, mean amplitude, median amplitude, mean frequency, median frequency, skewness, peak to peak-the difference between maximum and minimum peak, peak to rms, root sum of square (rssq) and variance. Using PCA on these 20 features, returned a matrix containing the features in order of their significance.

From this feature vector, the first 8 features were used. The first five, six or seven features give high percent accuracy too. However, they had outliers shown in the error histogram. Thus, using the first eight features avoided the occurrence of outliers and reduce overfitting. Hence, this model ensures a very high percentage of accuracy.

References

1. Abdi, H., Williams, L.J.: Principal component analysis. Wiley Interdisc. Rev. Comput. Stat. **2**(4), 433–459 (2010). doi:10.1002/wics.101
2. Asadi, S., Rao, C., Saikrishna, V.: A Comparative study of face recognition with principal component analysis and cross-correlation technique. Int. J. Comput. Appl. **10**(8), 17–21 (2010). doi:10.5120/1502-2019
3. Cybenko, G.: Continuous Valued Neural Networks with Two Hidden Layers are Sufficient, pp. 303–314 (1988)
4. Fong, S., Yang, X., Deb, S.: Swarm search for feature selection in classification. In: 2013 IEEE 16th International Conference on Computational Science and Engineering (2013). doi:10.1109/cse.2013.135
5. Funahashi, K.: On the approximate realization of continuous mappings by neural networks. Neural Networks **2**(3), 183–192 (1989). doi:10.1016/0893-6080(89)90003-8
6. Hagan, M.T., Demuth, H.B., Jesús, O.D.: An introduction to the use of neural networks in control systems. Int. J. Robust Nonlinear Control **12**(11), 959–985 (2002). doi:10.1002/rnc.727
7. Hornik, K., Stinchcombe, M., White, H.: Multilayer feedforward networks are universal approximators. Neural Networks **2**(5), 359–366 (1989). doi:10.1016/0893-6080(89)90020-8
8. Howard, W.: Pattern recognition and machine learning. Kybernetes **36**(2), 275 (2007). doi:10.1108/03684920710743466. i-xx, pp. 740. Springer, Heidelberg (2006). ISBN 0-387-31073-8, $74.95 Hardcover
9. Li, C., Diao, Y., Ma, H., Li, Y.: A statistical PCA method for face recognition. In: 2008 Second International Symposium on Intelligent Information Technology Application (2008). doi:10.1109/iita.2008.71
10. Hornik, K., Stinchcombe, M., White, H.: Multilayer feedforward networks are universal approximators. Neural Networks **2**(5), 359–366 (1989). doi:10.1016/0893-6080(89)90020-8
11. Mcculloch, W.S., Pitts, W.: A logical calculus of the ideas immanent in nervous activity. Bull. Math. Biophys. **5**(4), 115–133 (1943). doi:10.1007/bf02478259
12. Meruelo, A.C., Simpson, D.M., Veres, S.M., Newland, P.L.: Improved system identification using artificial neural networks and analysis of individual differences in responses of an identified neuron. Neural Networks **75**, 56–65 (2016). doi:10.1016/j.neunet.2015.12.002
13. Murali, M.: (2015). Principal component analysis based feature vector extraction. Indian J. Sci. Technol. (2015)
14. Phillips, P., Flynn, P., Scruggs, T., Bowyer, K., Chang, J., Hoffman, K., Worek, W.: Overview of the Face Recognition Grand Challenge. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2005) (2005). doi:10.1109/cvpr.2005.268
15. Rosenblatt, F.: The perceptron: a probabilistic model for information storage and organization in the brain. Psychol. Rev. **65**(6), 386–408 (1958). doi:10.1037/h0042519
16. Sakar, B.E., Isenkul, M., Sakar, C.O., Sertbas, A., Gurgen, F., Delil, S., Kursun, O.: Collection and analysis of a parkinson speech dataset with multiple types of sound recordings. IEEE J. Biomed. Health Inform. **17**(4), 828–834 (2013). doi:10.1109/jbhi.2013.2245674

17. Schmidhuber, J.: Deep learning in neural networks: an overview. Neural Networks **61**, 85–117 (2015). doi:10.1016/j.neunet.2014.09.003
18. Tamura, S., Tateishi, M.: Capabilities of a four-layered feedforward neural network: four layers versus three. IEEE Trans. Neural Networks **8**(2), 251–255 (1997). doi:10.1109/72.557662
19. Hori, T., Kubo, Y., Nakamura, A.: Real-time one-pass decoding with recurrent neural network language model for speech recognition. In: 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (2014)

Machine Learning-Based Method and Its Performance Analysis for Occupancy Detection in Indoor Environment

Sachin Kumar[1](✉), Shobha Rai[2], Rampal Singh[3], and Saibal K. Pal[4]

[1] Department of Computer Science and Cluster Innovation Centre,
University of Delhi, Delhi, India
sachin.blessed@gmail.com
[2] Cluster Innovation Centre, University of Delhi, Delhi, India
rai.shobha18@gmail.com
[3] DDUC, University of Delhi, Delhi, India
rprana@gmail.com
[4] Defence Research Development Organisation, Delhi, India
skptech@yahoo.com

Abstract. Occupancy detection is very interesting research problem which may help in understanding ambient dynamics of the environment, resource utilisation, energy conservation and consumption, electricity usages and patterns, security and privacy related aspects. In addition to this, achieving good accuracy for occupancy detection problem in the home and commercial buildings can help in cost reduction substantially. In this paper, we explain one experiment in which data for occupancy and ambient attributes have been collected. This paper develops machine learning-based intelligent occupancy detection model and compare the results with several machine learning techniques in a detailed manner.

Keywords: Occupancy detection · CART · Naive Bayes · SVM · Logistic regression · LDA · Classification

1 Introduction

Occupancy detection is an important research problem which has many applications in different domains. Mainly occupancy detection in residential and commercial buildings can help a great deal in understating so many things about the static and dynamic knowledge in the ambient environment. Some of the applications are energy conservation and monitoring of energy consumption, electricity and appliances use pattern, movement, and dynamics inside the buildings, energy cost analysis based on appliances, security and privacy related applications. There were conducted many research studies to predict the occupancy status in buildings. Such studies reported that such applications can save as per estimation from 30 to 42% energy [5,7,8]. Experiments and research conducted

S.M. Thampi et al. (eds.), *Advances in Signal Processing and Intelligent Recognition Systems*, Advances in Intelligent Systems and Computing 678,
DOI 10.1007/978-3-319-67934-1_21

by [2] has shown that energy saving was about 37% and varying from 29% to 80% [3] when occupancy related insights were used as an input for Heating Ventilation Air Conditioning (HVAC) control system. Machine learning approaches and intelligent techniques have also been used in many diverse and applied fields such as in health information systems [16], knowledge-based information systems [15], in intelligent and sustainable software development [20] and ICT-based social media discovery of consumer service delivery [13,21]. At present sensors based system can be used and developed easily to collect the data. In this research paper, occupancy detection is being predicted with the data based on environment obtained with the help of the sensors. Data contains attributes such as light, temperature, humidity, and CO2. Occupancy attribute has two values only occupied or not occupied. The good thing about the experiment is that it is based on environment and not much investment and infrastructure are required. As far as literature is concerned, this paper addresses occupancy as a classification problem with two classes occupied and not occupied. Several works have been done using some machine learning methods with different data sets using several approaches based on sensors, camera, videos monitoring and others. Each of them has their own advantages and disadvantages. [18] gave an improved method based on stochastic occupancy detection model. With the help of passive infrared detection, digital video cameras, and CO2 sensors [19], a model to detect occupancy was developed which is based on Bayesian statistics. This research reports that the model brings a reduction in average error of about 70% to 11%. Adding to this, [9] developed two models. The first model developed was based on multivariate Gaussian distribution technique of machine learning. In this experiment digital cameras were used for data collection. The second model developed was based on Agent-Based Model (ABM) technique of machine learning which can help in analysis and simulation of mobility patterns. Further, some researchers tried to detect a room's occupants count and developed a dynamic model with the help ventilation, temperature and Carbon dioxide data [6]. The occupancy figure of 88% was displayed by the developed model and this level is better than the one obtained through the models based on Neural Network estimators technique and Support Vector Machine (SVM) technique. To detect occupancy some models were also developed which use the data of wireless sensor network [8]. To determine occupants count this model uses cameras equipment. With this model, up to 42% energy can be saved annually along with the adequate amount of standards of comfort in the room. Extreme Learning Machine (ELM) has been used in classification and regression applications such as temperature forecasting, energy forecasting of buildings [12,14]. This work improves the previous research and development. To achieve better occupancy detection accuracy, we require monitoring equipment's which provides higher resolution and accuracy [4]. For predicting occupancy accurately with collected data we have used several machine learning methods and developed a model in this paper. This paper does an in-depth analysis of the performance and accuracy of several methods of machine learning to propose the most accurate methods for occupancy detection.

2 Model-Machine Learning Approaches

2.1 Linear Discriminate Analysis

Linear discriminant analysis (LDA) [17] is a effective machine learning techniques which is a generalized version of Fisher's linear discriminant [22]. Fisher's linear discriminate method is famously used in many domains such as statistics, pattern recognition, pattern analysis and machine learning for the purpose of the linear combination of input attributes and features that characterize or classify two or more classes of objects. After obtaining the resulting object, it can be taken like a linear classifier to reduce the dimensions in the experiment before classification.

2.2 Logistic Regression

It's a model based on regression in which on the basis of input features, one dependent variable or categorical is predicted. Categorical variable or class variable can be binary of having two or many classes. In this case, where the no of the dependent variable is representing more than two classes are called multinomial logistic regression. First Statistician David Cox developed the model of Logistic regression in 1958 [10]. In different subjects, different terminologies are used like in economics, qualitative discrete choice model's example is logistic regression. For instance, We need to predict the vote of a person based on certain parameters such as gender, age, and immigration so Mathematically it is defined as

$$Pr(Y=1|S,P,I) = \frac{exp(\beta_0 + \beta_1 Gender + \beta_2 Age + ...\beta_1 2 immigration)}{1 + exp(\beta_0 + \beta_1 Gender + \beta_2 Age + ...\beta_1 2 immigration)} \tag{1}$$

2.3 CART

In machine learning method, Decision Trees (DT) is one of the important types of algorithm for prediction and modelling. There are classical decision trees that have been developed and are being implemented but there are some modern variations of them such as a random forest that have shown most powerful performance in terms of accuracy. One of the decision tree algorithm in modern time is Classification And Regression Trees (CART) methodology introduced by Charles Stone, Richard Olshen, Jerome Friedman and Leo Breiman and is structured like question and answer set of a system. It asks a sequence of questions and answers to those question determine what can be the next question based on features and their relevance. The result of these questions and answers develop a tree like data structure with nodes. The nodes at the ends are called terminal nodes where question ends. CART uses entropy as information gain mechanism and creates Binary trees which are two children tree with splitting criteria is performance based on best split point.

$$\Phi(s/t) = 2P_L P_R \sum_{j=1}^{m} |P(C_j|t_L) - P(C_j|t_R)| \tag{2}$$

In which R and L indicates the right and left subtrees of the current node where decision is being made based on information gain. P_R, P_L subtrees probability (tuples in left or right sub-tree)/(tuples in training set).

2.4 KNNs

In intelligent systems development, there are domains related to problems of pattern recognition and machine learning. In machine learning domain,for classification and regression work a non-parametric method the k-nearest neighbor's algorithm (k-NN) approach is used [1]. Suppose a feature space is with N dimension is given and then task associated can be classification and regression. The predicted class variable is category and output variable is class in the k-NN classification method. In which our purpose is to identify the object based on feature variables which class it belongs to. The KNNs follows the approach in which an object is classified on the basis of maximum number of vote of the neighbors. This is for the k nearest neighbours (k being positive integer) of the object assigned to the class. The object can be allocated to the class having single nearest neighbour if k value is equal to 1. But in k-NN regression approach, the object's property value is the output (a real number) which is an average of k nearest neighbor's values. Suppose there are following training examples $\langle x_i, f(x_i)\rangle$ where we want to implement KNNs. Given x_q query instance, first of all locate x_n nearest training example, then get an estimate for $\hat{f}(x_q)f(x_n)$ k-Nearest Neighbor: is given x_q, for real-valued take mean of f values of the k nearest neighbours and for discrete-valued target function take vote of its k nearest neighbours

$$\hat{f}(x_q)\frac{\sum_{i=1}^{k} f(x_i)}{k} \tag{3}$$

Most of the algorithms have their special properties which make them best candidates for the special type of problems. This is also true for KNNs.The KNNs is well suited for cases where instances map to points in $\Re^n$ are less than 20 attributes per instance and there are lots of training data. There are some benefits also of using KNNs such as no data or information lose, complex target functions learning and very fast training. Apart from this there are disadvantages also, irrelevant attributes can fool easily and query time is slow.

2.5 Support Vector Machine

One of the important machine learning algorithms Support Vector Machine (SVM) is based on discriminate classifier or separating hyperplane. SVM is trained with given labelled training data in supervised learning and the outputs of SVM is an optimal hyperplane that classify some new and unique examples by separating them from one another. The mathematical formulation is given below. SVM has the constraint

$$\sum_{i} y_i \alpha_i = 0, \tag{4}$$

which makes the total weight for the positive class equal to that of the negative class. Suppose, we are given data $(x_i, y_i), \quad x \in R^d, y \in \{-1, 1\}$. We want a linear classifier in an infinite-dimensional kernel space,

$$g(x) = sign(\phi(w) \cdot \phi(x) + b), \tag{5}$$

where

$$\phi(w) \cdot \phi(x) = K(w, x). \tag{6}$$

The SVM optimization is $\phi^*(w) = \arg\min_{\phi(w)} \frac{1}{2}\phi(w)^2$, *such that* $y_i(\phi(w) \cdot \phi(x_i) + b) \geq 1$. So of all the classifiers which correctly classify the data, we want the one closest to the origin. From a Bayesian perspective, this is choosing the most probable classifier under a zero mean normal prior, which has likelihood above a certain threshold. The Lagrangian is

$$L(\phi(w), b, \alpha) = \frac{1}{2}||\phi(w)||^2 - \sum_i \alpha_i(y_i(\phi(w) \cdot \phi(x_i) + b) - 1). \tag{7}$$

The stationary conditions are $L(\phi(w), b, \alpha)\phi(w) = \phi(w) - \sum_i y_i\alpha_i\phi(x_i) = 0$, $L(\phi(w), b, \alpha)b = \sum_i y_i\alpha_i = 0$. So the weight vector is a linear combination of the data points:

$$\phi(w) = \sum_i y_i\alpha_i\phi(x_i). \tag{8}$$

The classifier is then $g(x) = \text{sign}\Big(\sum_i y_i\alpha_i\phi(x_i) \cdot \phi(x) + b\Big)$ $= sign\Big(\sum_i y_i\alpha_i K(x_i, x) + b\Big)$.

2.6 Naive Bayes

Naive Bayes classifiers belongs to the probabilistic classifiers family which fundamentally apply along with independent assumptions among the features "Bayes' theorem". Naive Bayes has been known and famous for many decades and still remains a competitive and popular one of the area with some advance methodologies covering support vector machine. Its applicability and performance are appreciated many domains. The good thing about Naive Bayes classifiers in a learning problem is that they are highly scalable and requires a number of parameters form of linearity. In this approach, for achieving Maximum-likelihood training, the evaluation of a closed-form expression with linear running time complexity is done rather than expensive iterative approximation that. Mathematical formulation of Naive Based Classifier with equation and posterior probability is given as following $P(\omega_j|x) = \frac{p(x|\omega_j) \cdot P(\omega_j)}{p(x)}$ $posterior\,probability = \frac{likelihood \cdot prior\,probability}{evidence}$ One the probabilities have been calculated, in addition to this, required objective function is given as $g_1(x) = P(\omega_1|\ x), \quad g_2(x) = P(\omega_2|\ x), \quad g_3(x) = P(\omega_2|\ x)$.

2.7 Gradient Boost Machine

Gradient boosting has three elements. One is the loss function. second is a weak learner and this is an additive model that adds weak learners for minimizing the weak learning. The function understanding used depends on the problems itself. The basic property of loss function is that it must be differentiable. For instance, in regression task, we can use a squared error and in the classification task, the logarithmic loss can be used. Decision trees method is basically used in gradient boosting as the weak learner. After the weak learners, trees are incorporated one at a time and existing trees in the model are not changed. One of the important procedure of the gradient descent procedure is to maximize the accuracy and minimize the loss. In adding tree procedure, the output for the newly added tree is then added to the output of the existing set of already added trees in an order to improve the final output. The development of boosting algorithms in many domain of machine learning and statistics apart from regression and classification has been driven by this functional gradient view of boosting.

3 Methodology

3.1 Data Description

Occupancy data set has been obtained from an experiment that stated the binary classification for room occupancy. The data contain input feature space with attributes such as temperature, humidity, light and CO2 and time stamp. Dataset is multivariate with no of instance 20560 and number of attributes equal to 7. The data have been obtained from UCI machine learning repository [11].

3.2 Model and Its Description

This paper analyses the performance of the model based on many states of art machine learning models to predict the occupancy in the room whose data has been collected from the sensors. Proposed model first set the sensors and obtain the data from all required sensors which provide the data with features such as date, temperature, relative Humidity, light, CO2, humidity ratio and occupancy as stated on [11].

This data is then imported into the set of experiment and data cleaning process is performed if any value is missing and other rectification of data as displayed in Fig. 1. After this process, data is sent to statistical analysis phase which performs data and features analysis stating the importance of the features and correlation with the occupancy attributes as shown in Fig. 3 and Table 1. Here Pearson correlation is also determined in order to know about the relation and their quantification. After statistical analysis phase, we have enough insights about the data and feature space which is then used to feature selection and feature generation. In our experiment, timestamp related information is also used. Once the proper feature space is developed, it is passed into machine learning models based on K-Nearest neighbor (KNN), Linear discriminate analysis

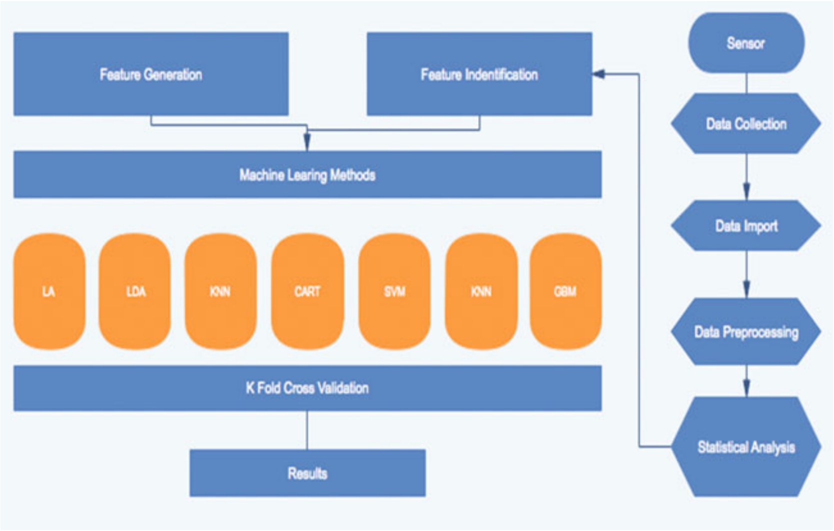


Fig. 1. Performance analysis model

Table 1. Feature ranking details

Serial no.	Feature no.	Ranking coefficient	Feature name
1	Feature 2	0.470111	Temperature
2	Feature 3	0.301546	Humidity
3	Feature 0	0.124318	Light
4	Feature 4	0.05819	CO2
5	Feature 1	0.045835	Humidity ratio

(LDA), Linear regression (LR), Support vector machine (SVM), Naive Bayes (NB), Gradient boosting machine (GBM). Feature space is passed to each of the machine learning model and accuracy is evaluated as stated in the evaluation section. after this to obtain the generalist performance k cross-fold validation system in applied and values of k equals to 10. this gives good results. This model is run with two type of data one without the time stamp data and other with added features with the time stamp and related features. When time stamp or date related information is incorporated into the model it performs better.

3.3 Performance Evaluation

The task associated with the data is classifications which depend on the count of the correctly classified items. There are four possibilities of the count of classification results that can be counted in each case of classification. For evaluation of our model, we use accuracy defined as follows.

$$Accuracy = (TP + TN)/(TP + TN + FP + FN) \tag{9}$$

Where TP, TN, FP and FN is True Positive, True Negative, False Positive and False Negative respectively. Obtaining good level of accuracy in a generalized environment is desirable for machine learning model. For this purpose, in this experiment cross-validation technique has been used for evaluating predictive models with partition of data points between testing and training. Training instances are used to train the model while testing is used to validate the performance in generalized environment. For generalized performance, we have selected k-fold cross-validation in which we do random partitioning of original set into k equal size subset. Out of these k equal size subset, for testing the model one subset is used as the validation data and for training rest of the k-1 subsets are utilized. Then this process of cross-validation is repeated k number of times and we call it as folds where each and every k subset used once only in terms of validation data. Now for producing single estimation take an average of k results come out of such systems. This is the testing of generalized ability of the model. In our experiment, we have selected k value equals to 10 hence it is 10 fold cross validation.

4 Results and Discussion

This paper has taken a study on the development of the model based on latest machine learning techniques to predict the occupancy. In this discussion, We first took the data and move towards the implementation as discussed in the methodology section.

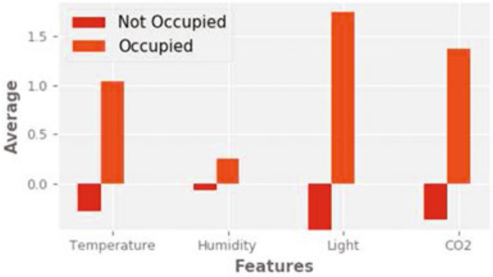


Fig. 2. Occupancy display plot

This data is imported and cleaned with NAN values. Once this part is complete, data is passed to the statistical phase where feature correlation among themselves, with target attribute and feature ranking, feature selection is done as displayed in Figs. 2 and 3. This way some attribute already given are expanded. Model add the feature of time and data specifically in the training and testing which produces good results. Results obtained from the experiment are displayed

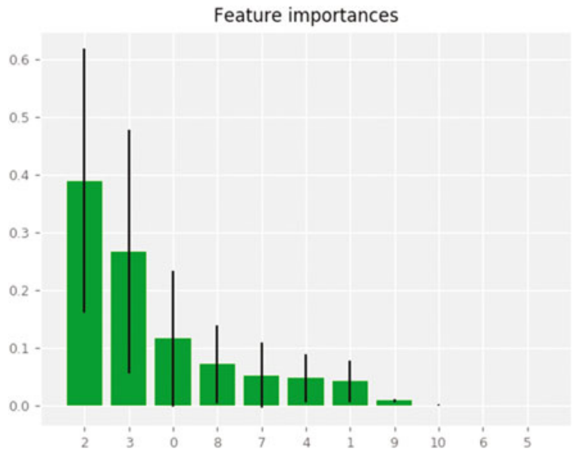


Fig. 3. Feature importance plot

in Tables 2 and 3. Where Table 2 describes results without adding time and date related information and applying the proposed model and doing the comparative study with machine learning techniques such as LR, LDA, KNN, CART, NB, SVM and GBM. Accuracy varies from lowest 58.81% to maximum of 98.82%. The model produces lowest accuracy for SVM and highest for the LR. If extra information processed from date and time with existing feature space is added, then model produces almost same results but there is the remarkable improvement in the model accuracy by SVM which reaches from 58.81% to 78.77%. The second model shows improved performance in CART and SVM otherwise the previous model performed well.

Table 2. Performance analysis without improved feature space

Machine learning	Accuracy	Standard deviation
Logistic regression	0.988218	0.023488
Linear discriminate analysis	0.960576	0.057737
K-Nearest neighbour	0.971022	0.037281
Classification and regression tree	0.931589	0.128586
Naive Bayes	0.964753	0.056778
Support vector machine	0.588199	0.239127
Gradient boosting machine	0.970407	0.046774

While displaying the accuracy comparatively, it is observed that logistic regression performed better than any other machine learning model. Then comes the KNN with the accuracy of 97.10% followed by GBM. After that accuracy

Table 3. Performance analysis with improved feature space

Machine learning	Accuracy	Standard deviation
Logistic regression	0.970285	0.034031
Linear discriminate analysis	0.961559	0.045556
K-Nearest neighbour	0.976547	0.034958
Classification and regression tree	0.924953	0.133456
Naive Bayes	0.964016	0.058992
Support vector machine	0.787701	0.269817
Gradient boosting machine	0.968073	0.043683

comes to 96% range for LDA and NB. This way the generalised performance of the on this data set from machine learning methods is good for LR and KNN. This proposed model displays the comparative study of the various machine learning methods in Fig. 4 in pictorially. Hence in this study, we can say that with added attribute SVM accuracy improved significantly and logistic regression performed well for the generalised accuracy with k = 10 fold cross-validation model.

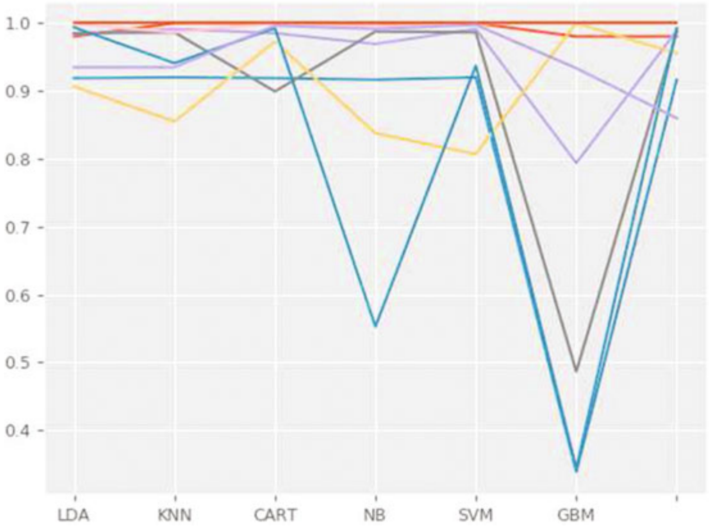


Fig. 4. Performance comparison of machine learning methods

5 Conclusion

This paper discusses the experiment based on sensor data to predict the occupancy in the room.Sensor data have many advantages as it requires less computational power as compared to image processing or videos processing for occupancy detection. This paper developed model based on machine learning methods and evaluated and compared their performance among themselves in which simple linear regression models performed better. Linear regression and KNNs based model produced the maximum accuracy of order 97.68% and this accuracy is generalised accuracy obtained with 10 fold cross validation with variation quite low of order 0.3498. Improvement in this model is that, the model incorporate time stamp details also as input part when the prediction is done which has improved accuracy for different models but a significantly improve accuracy for SVM which goes from 68% to 78.77%. Future work could be to use advanced learning such and Deep learning and applying feature generation techniques to add more insight of the data into training process.

References

1. Altman, N.S.: An introduction to kernel and nearest-neighbor nonparametric regression. Am. Stat. **46**(3), 175–185 (1992)
2. Brooks, J., Goyal, S., Subramany, R., Lin, Y., Middelkoop, T., Arpan, L., Carloni, L., Barooah, P.: An experimental investigation of occupancy-based energy-efficient control of commercial building indoor climate. In: 53rd IEEE Conference on Decision and Control, pp. 5680–5685. IEEE (2014)
3. Brooks, J., Kumar, S., Goyal, S., Subramany, R., Barooah, P.: Energy-efficient control of under-actuated HVAC zones in commercial buildings. Energy Build. **93**, 160–168 (2015)
4. Candanedo, L.M., Feldheim, V.: Accurate occupancy detection of an office room from light, temperature, humidity and CO2 measurements using statistical learning models. Energy Build. **112**, 28–39 (2016)
5. Dong, B.: Sensor-based occupancy behavioral pattern recognition for energy and comfort management in intelligent buildings (2009)
6. Ebadat, A., Bottegal, G., Varagnolo, D., Wahlberg, B., Johansson, K.H.: Estimation of building occupancy levels through environmental signals deconvolution. In: Proceedings of the 5th ACM Workshop on Embedded Systems for Energy-Efficient Buildings, BuildSys 2013, pp. 8:1–8:8. ACM, New York (2013)
7. Carreira-Perpin, M., Erickson, V.L., Cerpa, A.E.: Observe: occupancy-based system for efficient reduction of HVAC energy. In: 10th Information Processing in Sensor Networks (IPSN). IEEE (2011)
8. Erickson, V.L., Carreira-Perpiñán, M.Á., Cerpa, A.E.: Occupancy modeling and prediction for building energy management. ACM Trans. Sen. Netw. **10**(3), 42:1–42:28 (2014)
9. Erickson, V.L., Lin, Y., Kamthe, A., Brahme, R., Surana, A., Cerpa, A.E., Sohn, M.D., Narayanan, S.: Energy efficient building environment control strategies using real-time occupancy measurements. In: Proceedings of the First ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Buildings, BuildSys 2009, pp. 19–24. ACM, New York (2009)

10. Freedman, D.: Statistical Models: Theory and Practice. Cambridge University Press, Cambridge (2009)
11. Lichman, M., Bache, K.: UCI machine learning repository. University of California (2013)
12. Kumar, S., Rai, S., Singh, R., Pal, S.K.: ELM variants comparison on applications of time series data forecasting. In: 2016 International Conference on Advances in Computing, Communications and Informatics (ICACCI), pp. 1404–1409, September 2016
13. Kumar, S., Pal, S.K.: Empirically developed integrated ICT framework for PDS in developing countries. In: 2013 Third World Congress on Information and Communication Technologies (WICT), vol. 3, pp. 235–240. IEEE (2013). ISBN: 978-1-4799-3230-6/13
14. Kumar, S., Pal, S.K., Singh, R.P.: Intelligent energy conservation: indoor temperature forecasting with extreme learning machine, pp. 977–988. Springer International Publishing, Cham (2016)
15. Kumar, S., Shobha, R., Pal, S.K.: A new sustainable prototype USP for education information system. In: International Conference on Futuristic Trends in Computational Analysis and Knowledge Management, vol. 1. IEEE (2015). ISBN: 978-1-4799-8432-9
16. Kumar, S., Singh, R., Pal, S.K.: A conceptual architectural design for intelligent health information system: Case study on India. In: Proceedings of 7th International Conference On Quality, Reliability, Infocom Technology and Business Operations (Trends And Future Directions). Springer (2017)
17. Martinez, A.M., Kak, A.C.: PCA versus lDA. IEEE Trans. Pattern Anal. Mach. Intell. **23**(2), 228–233 (2001)
18. Richardson, I., Thomson, M., Infield, D.: A high-resolution domestic building occupancy model for energy demand simulations. Energy Build. **40**(8), 1560–1566 (2008)
19. Lin, Y., Oggianu, S.M., Narayanan, S., Frewen, T.A., Meyn, S., Surana, A.: A sensor-utility-network method for estimation of occupancy in buildings. In: Proceedings of the 48th IEEE Conference on Decision and Control (CDC 2009) and 28th Chinese Control Conference CCC 2009, Shanghai, P.R. China, pp. 1494–1500. IEEE (2009)
20. Kumar, S., Pal, S.K.: An approach to ensure superior and sustainable software development performance. In: International Conference on Computing for Sustainable Global Development, pp. 115–117. IEEE Explorer Conference Id 32410 (2014). ISBN: 978-93-80544-10-6
21. Kumar, S., Pal, S.K.: ICT integrated social media framework for consumer awareness in society using ICT tools. In: IEEE/ACIS 13th International Conference on Computer and Information Science (ICIS), vol. 13, pp. 229–233. IEEE (2014)
22. Welling, M.: Fisher linear discriminant analysis. IEEE Trans. Pattern Anal. Mach. Intell

Identifying Issues in Estimating Parameters from Speech Under Lombard Effect

M. Aiswarya(✉), D. Pravena, and D. Govind

Centre for Computational Engineering and Networking (CEN),
Amrita School of Engineering, Amrita Vishwa Vidyapeetham,
Coimbatore 641112, Tamilnadu, India
aishmadhu.92@gmail.com, d.pravena@gmail.com, d_govind@cb.amrita.edu
http://www.amrita.edu/campus/coimbatore

Abstract. Lombard effect (LE) is the phenomena in which a person tends to speak louder in the presence of loud noise, due to the obstruction of self-auditory feedback. The main objective of this work is to develop a dataset for the study of LE on speech parameters. The proposed dataset comprising of 230 utterances each from 10 speakers, consists of the simultaneous recording of speech and ElectroGlottoGram (EGG) of speech under LE as well as neutral speech recorded in a noise free condition. The speech under LE is recorded at 5 different levels (30 dB, 15 dB, 5 dB, 0 dB and −20 dB) of babble noise. The level of LE in the developed dataset is demonstrated by comparing (a) the source parameters, (b) speaker recognition rates and (c) epoch extraction performance. For the comparison of source parameters like pitch and Strength of Excitation (SoE), the neutral speech and speech under LE are compared. Based on the comparison, high pitch and low SoE are observed for the speech under LE. Also, lower recognition performance is observed when a Mel Frequency Cepstral Coefficient (MFCC) - Gaussian Mixture Model (GMM) based speaker recognition system built using the neutral speech, is tested with the speech under LE obtained from the same set of speakers. Finally, on the basis of the comparison of epoch extraction from neutral speech and speech under LE, the utterances with LE is observed to have higher epoch deviation than that for neutral speech. All these experiments confirm the level of LE in the prepared database and also reinforces the issues in processing the speech under LE, for different speech processing tasks.

1 Introduction

The human speech production varies with the environment in which the speech is produced [1], on one's state of mind or emotions [2]. The ultimate aim of the speech production system is to convey the information in the best possible form. Lombard effect (LE) is the phenomena in which a person tends to speak louder in the presence of loud noise, due to the obstruction of self-auditory feedback [1]. LE is an involuntary effort taken by the speaker to convey a clear information. For example, when you talk to a person who is listening to some loud music

S.M. Thampi et al. (eds.), *Advances in Signal Processing and Intelligent Recognition Systems*, Advances in Intelligent Systems and Computing 678,
DOI 10.1007/978-3-319-67934-1_22

through earphones, that person replies in a louder voice than usual. The articulatory and acoustic parameters will be adjusted to increase the intelligibility of speech in the presence of noise [1], which makes the features of speech under LE different from that of normal speech [3]. Though LE in speech helps to increase the intelligibility of the speech produced, it is an undesirable phenomenon with respect to a speaker recognition system.

An automatic speaker recognition system uses certain features of a person's speech to correctly identify the speaker [3]. Since the features of neutral speech vary from the features of speech under stress, LE downgrades the performance of a speaker recognition system [3]. A system may give good performance with improved training but fails if there is even a slight change in the test data [4]. In spite of various technical changes made in the speaker recognition systems to increase its robustness, a speaker recognition system still faces many challenges and issues that degrade the performance of system [5]. Bapineedu in [6] has observed that the LE is reflected mainly on the features of the excitation source. Features like instantaneous fundamental frequency, duration of voiced region and strength of excitation can be analysed to find the difference between a normal speech and speech under LE. Bapineedu et al. in [1] have analysed speech under LE with different types and levels of noise.

Figure 1 shows the speech signal of neutral speech (a), and speech under LE (b) and their corresponding narrowband spectrograms (c) and (d), for the same sentence 'He would be more popular'. Although both the spectrograms are for the same sentence recorded by the same speaker, there is a significant difference between the two signals due to LE. The fundamental frequency and its harmonics are present in the vocal fold vibrations [7]. The first horizontal band represents the fundamental frequency and the successive horizontal bands represent each harmonic. The source parameter pitch is correlated with the fundamental frequency and width of the horizontal striations of a narrow band spectrum. From the Fig. 1 it can be seen that the width of the first horizontal striation which indicates the fundamental frequency, of the Lombard speech is more than that for the neutral speech. All the successive bands of the spectrogram of Lombard speech are also wider than that of the bands of the spectrogram of neutral speech. This implies that the pitch of a speech under LE is greater than that of

Table 1. Existing datasets for the study of LE

Dataset	No. of speakers	No. of sentences/words	Noise	Datatype
UT-Scope [8] (Speech under Cognitive and Physical Stress and Emotion)	59	100 Sentences (TIMIT corpus) + 5 tokens of digits (0–9)	(a) Highway (b) Large crowd (c) Pink noise	Speech
SUSAS [9] (Speech Under Simulated and Actual Stress)	9 (All male)	70 utterances	Noise (85 dB SPL)	Speech
CLSD [10] (Czech Lombard Speech Database)	26 (12 female,14 male)	108 utterances	(a) Car noise (b) artificial band noises	Speech

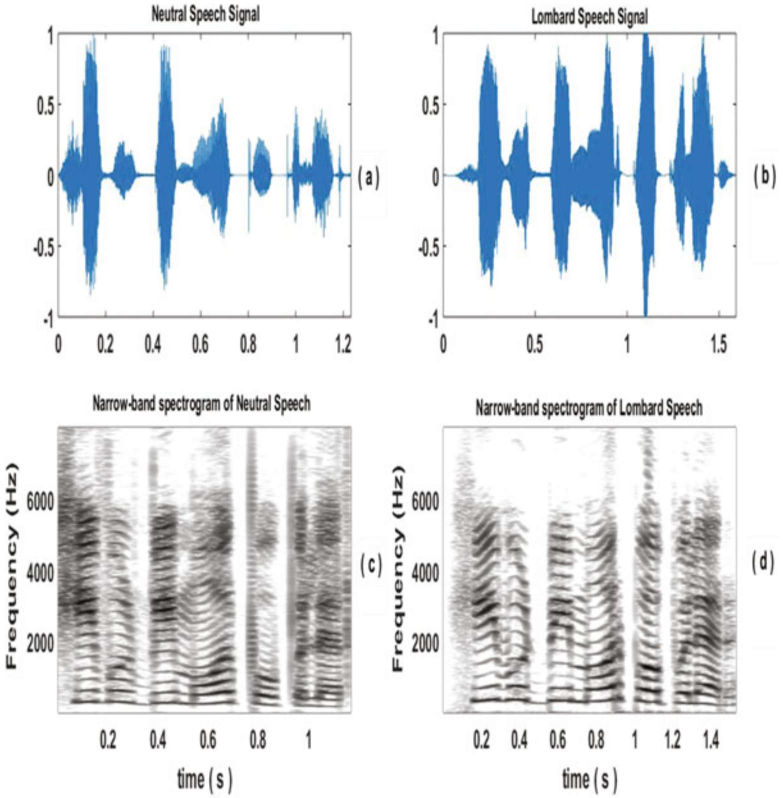


Fig. 1. Comparison of characteristics of neutral speech and speech under LE: (a) speech signal of neutral speech, (b) speech under LE, (c) and (d): their corresponding spectrograms

neutral speech. These changes in source parameters lead to poor performance of a speaker recognition system.

In order to solve the problem of speaker recognition system, a comparative study of neutral speech and speech under LE is necessary. Many studies have been carried out on speech under LE with the help of conventional datasets. Table 1 shows the details of some of the existing datasets for the study of speech under LE. The SUSAS dataset is the most commonly used dataset [10] to study the speech production in a stressed condition and to find the impact of LE on speech systems. SUSAS dataset is a small database partly dedicated to LE and is publicly available [10]. There are several speaking styles in this database like the normal, slow, clear, Lombard etc. The recordings of 9 speakers (all male) with 70 utterances by each speaker are present in the dataset. 35 commonly used words in aircraft, each repeated once constitute the 70 utterances. The Lombard speech is recorded at a noise level of 85 dB SPL. The clear speech corresponding to the speech under LE, of the 9 speakers is also included in the dataset [9].

Hansen et al. in [9] have observed that speech recognition performance has degraded significantly with the introduction of stress. The UT-Scope [8] is a comparatively larger database with speech under LE of 59 speakers recorded at three types and levels of noise. Except for UT-Scope, all other datasets have used a maximum of only two noise types and levels. The CLSD [10] is a dataset in the native language. All these datasets consist of only speech data and only a few of them are publicly available.

According to Bapineedu et al. in [1], the environment in which the speech is produced, speaker, the nature and extent of noise that causes LE are the main factors that influence the speech under LE. So, the development of a larger database when compared to the publicly available SUSAS dataset, with respect to the number of speakers, utterances and noise levels, can benefit the study of speech under LE. The proposed dataset, which has speech under LE at five different noise levels, has to be validated against a standard dataset by analysing the speaker recognition performance and source parameters, in order to confirm the reliability. Since the LE is reflected mainly on the excitation source features [6], an analysis of the excitation source features under LE is required. The excitation source parameters can be easily analysed from the EGG signal [11], which is present in the proposed dataset. Features like instantaneous fundamental frequency and Strength of Excitation (SoE) can be analysed to find how speech under LE is different from normal speech and can be obtained with the knowledge of epoch locations. The performance analysis of ZFF algorithm has to be performed to identify the issues in estimating source parameters under LE.

The remaining sections are organised as follows. Section 2 explains the development of the proposed dataset for speech under LE. Section 3 explains the analysis of excitation source parameters under LE. Section 4 in two subsections describes the speaker recognition experiments performed on two different datasets. The performance analysis of epoch location extraction using ZFF is described in Sect. 5. The work is concluded in Sect. 6.

2 Development of Database for Speech Under LE

In the proposed method of data collection of speech under LE, 230 English sentences are selected from Proverbs and blogs as recording prompt. 10 students (6 female, 4 male) in the age-group of 22–25 years are involved in the collection of data. None of the 10 speakers reported having any speech or hearing impairment. Both speech and EGG signals are recorded simultaneously on a dual channel with 48kHz as the sampling rate. The EGG signals are recorded using the ElectroGlottoGraph (EGG) device. The speech under LE is recorded in simulated noisy condition, whereas the neutral speech is recorded in a noise-free environment for a comparative study. To record the speech under LE, the speakers are presented with babble noise through earphones and are asked to read out the sentences. The method used to record the speech under LE is taken from the work done by Sumitra et al. in [12]. Speech under LE at five different levels of babble noise (30 dB, 15 dB, 5 dB, 0 dB and −20 dB) are included in the proposed dataset. A total of 13,800 sentences (5 noises and 1 clear × 10 Speakers × 230 Sentences) are collected from the speakers. The tool (WaveSurfer) takes input from a microphone and it can be saved as .wav files. Some advantages of the proposed dataset over the existing datasets for the study of LE on speech parameters is that the developed dataset is a comparatively large database with speech recordings of continuous sentences at five different noise levels. Corresponding EGG signals of speech under LE are also available from which the source parameters can be easily analysed. In the present work, EGG signals with manually marked GCIs are used as ground truth for performance analysis of epoch extraction using ZFF.

3 Analysis of Excitation Source Parameters Under LE

The level of LE on the proposed dataset is confirmed by measuring the pitch and SoE of the speech signal. The average of the pitch (mean F0) and SoE of 50 files, from neutral speech and speech under LE at significant noise levels (0 dB and −20 dB) are evaluated to measure the quality and amount of LE caught in the database. The average values obtained for neutral speech and speech under LE, at two noise levels (0 dB and −20 dB) are shown in Table 2. Bapineedu et al. in [1] have observed that the mean F0 of speech under LE will be greater than the mean F0 for neutral speech, whereas the strength of excitation will be less for speech under LE when compared with the neutral speech [1]. From the table, it can be observed that the proposed dataset follows the same trend as in literature [1]. The mean F0 for neutral speech is less than that for speech under LE, whereas the SoE is more.

Table 2. Mean F0 and SoE values of speech signal of the proposed dataset

Noise level	Mean F0	Avg SoE
Neutral	206.76	0.401
0 dB	237.62	0.387
−20 dB	252.12	0.389

4 Effect of LE on Speaker Recognition System

Speaker recognition experiment is performed for the standard SUSAS dataset and the proposed dataset. Cepstral features like the Mel Frequency Cepstral Coefficients (MFCC), with a combination of Gaussian Mixture Model (GMM) classifier, is most commonly used in a speaker recognition system. The feature vectors of a speaker are modelled as Gaussian densities in GMM for speaker recognition systems [13]. Here, for the SUSAS dataset and the proposed dataset, a GMM based speaker recognition system is developed which is trained with neutral speech. 39 MFCC coefficients extracted for speech signal using the HTK toolkit is used as the feature in both cases [14]. The training files are subjected to a GMM model to obtain the best fit curve for the model and generate corresponding GMMs for each speaker class. The frame size and frame shift used are 20ms and 10ms respectively. Only the neutral speech and speech under LE at 0 dB and −20 dB noise levels, of the proposed dataset are considered for speaker recognition experiment. Two kinds of speaker recognition are performed for both the datasets. In the first case, both training and testing are done with neutral files whereas in the second case, training is with neutral speech and testing is with speech under LE.

Speaker Recognition System for SUSAS Dataset. An MFCC-GMM based speaker recognition system is developed for SUSAS dataset. Out of the 70 clear speech files, 60 files are taken as training data. The system is first tested with remaining 10 files of the clear speech. The same system is then tested with the corresponding 10 files of the speech under LE. The speaker recognition rates for the SUSAS dataset, obtained for 8 different Gaussian components is given in Table 3.

Table 3. Speaker recognition rates for the SUSAS dataset

No. of Gaussians	Accuracy (%)	
	Tested with 10 clear files	Tested with 10 Lombard files
8	81.11	42.22
16	87.78	44.44
32	90.00	90.00
64	88.89	40.00
128	83.33	41.11
256	73.33	30.00
512	61.11	26.67
1024	37.78	15.56

In the case of both training and testing with neutral speech, the maximum accuracy obtained is 90% and the minimum is 37.78%. The lowest accuracy is obtained for 1024 Gaussian mixtures. For the same number of Gaussian mixtures, when tested for speech under LE, the accuracy obtained is just 15.56%. Out of the 8 Gaussian models, 7 Gaussian models is found to have lower accuracy when tested with speech under LE. The degradation of speaker recognition system can be observed from the results obtained.

Table 4. Speaker recognition rates for the proposed dataset

No. of Gaussians	Accuracy (%)		
	Neutral	Lombard speech at 0 dB	Lombard speech at −20 dB
8	99.75	83.50	67.25
16	100.00	86.25	70.75
32	99.75	90.25	72.00
64	100.00	90.75	69.25
128	100.00	95.00	71.75
256	100.00	96.00	75.25
512	100.00	95.25	75.5
1024	100.00	95.25	77.00

Speaker Recognition System for the Proposed Dataset. The MFCC-GMM based speaker recognition system for the proposed dataset uses 190 files out of the 230 sentences of neutral speech as the training data. The remaining 40 files of neutral speech are used for testing. The system is again tested with the corresponding 40 files of speech under LE, at two significant noise levels (0 dB and −20 dB). The speaker recognition accuracies obtained for the proposed dataset for 8 different Gaussian components, is given in Table 4. Nearly 100% accuracy is obtained for all Gaussian models when tested with 40 neutral speech of all speakers. The same system when tested with the corresponding 40 sentences of the speech under LE at 0 dB noise, a degradation in the recognition rate is observed. The highest accuracy obtained is 96% and the lowest 83.50%. The highest accuracy obtained is again degraded to 77% when tested with corresponding 40 files of speech under LE at −20 dB noise. The lowest accuracy obtained for this case is 67.25%. A performance degradation of speaker recognition system is observed for speech under LE of the proposed dataset as observed for the standard SUSAS dataset.

5 Effect of LE on Epoch Extraction

Since the LE is reflected mainly in the excitation source features, a better understanding is possible with the analysis of the source parameters. The excitation source parameters like pitch and SoE can be easily analysed from the EGG signal [11] with the knowledge of the epoch locations. Pitch frequency is given by the inverse of the time interval between two successive epoch locations. SoE is the amplitude at the GCIs of differentiated EGG (DEGG). Among the various existing methods for epoch location extraction, ZFF (Zero Frequency Filtering) is the one method which gives a better performance. For this method to give an accurate result, significant energy is required around the impulse at zero frequency, otherwise, resulting in an inaccurate epoch location extraction [15]. The instant of glottal closure is usually the instant of maximum excitation and is termed as Glottal Closure Instant (GCI). The source parameters obtained using the epoch extraction varies with emotions and stress. The large pitch variations make the epoch extraction a difficult task [16].

The performance of the epoch extraction is measured in terms of Identification Rate (IDR), Missing Rate (MR), False Alarm Rate (FAR) and IDentification Accuracy (IDA). If e_t , e_{t-1}, $and\, e_{t+1}$ are the current, previous and succeeding epoch locations in the reference data respectively, then the larynx cycle is in the region (1/2)(e$_{t-1}$+e_t) < n < (1/2)(e_{t+1}+e_t). According to the number and locations of the epochs extracted, the four parameters [17] are described as

Identification Rate (IDR): The number of larynx cycles where exactly one epoch location is present, converted to percentage gives the IDR.

Miss Rate (MR): The number of larynx cycles where no epochs are identified, converted to percentage gives the MR.

Identification Error: Identification error indicates the timing error or deviation of the extracted epoch, from the epoch in the reference data, in the case where exactly one epoch is identified in a larynx cycle.

Identification Accuracy (IDA): IDA is given by the standard deviation of identification error. Lesser the value of IDA, greater the accuracy.

Performance analysis of epoch extraction using ZFF is carried out on a subset of the proposed dataset that includes 50 EGG and speech files from the neutral and speech under LE at two distinct noise levels (0 dB and −20 dB). The epoch locations of the DEGG signal are manually marked and are taken as the reference [15]. By comparing the epoch locations extracted using ZFF, with the reference, the four parameters (IDR, MR, FAR, IDA) are calculated. The results obtained for the performance analysis of epoch extraction using ZFF is given in Table 5 for EGG and Table 6 for speech respectively.

Table 5. Performance analysis for EGG

Noise level	IDR (%)	MR (%)	FAR (%)	IDA (ms)
Neutral	99.65	0.29	0.05	0.19
0 dB	99.63	0.34	0.03	0.23
−20 dB	99.63	0.33	0.03	0.22

Table 6. Performance analysis for speech

Noise level	IDR (%)	MR (%)	FAR (%)	IDA (ms)
Neutral	99.78	0.06	0.15	0.24
0 dB	99.08	0.11	0.81	0.3
−20 dB	97.36	0.21	2.4	0.33

In both the cases, a decrease in the IDR is observed for the speech under LE. The incorrect epoch extraction case can be further considered as either a missing or multiple epoch detections in a larynx cycle. A measure of these two is given by MR and FAR. The epoch extraction performance is found to be poor for the speech under LE, when compared to the neutral speech. Further, from both the tables, it can be seen that the IDA is more for the speech under LE when compared with neutral speech. Lesser the value of IDA implies a greater accuracy.

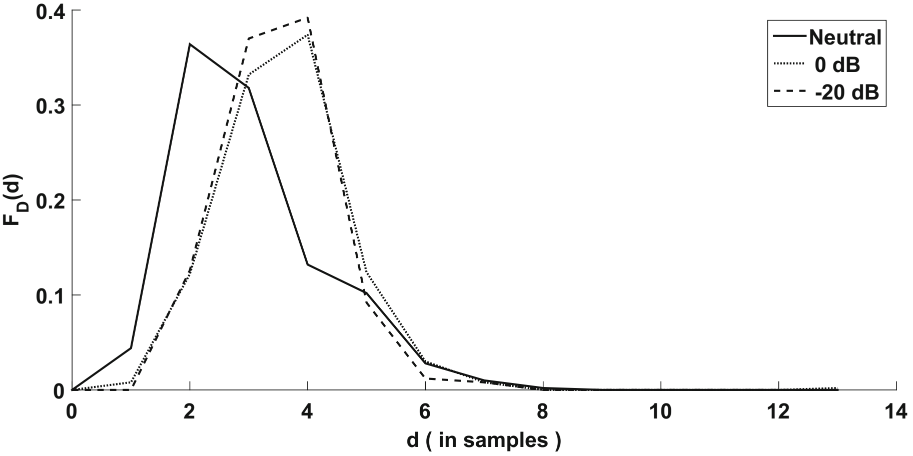


Fig. 2. Probability distribution of identification accuracies of EGG signals of Neutral speech, Lombard speech at 0 dB and Lombard speech at −20 dB

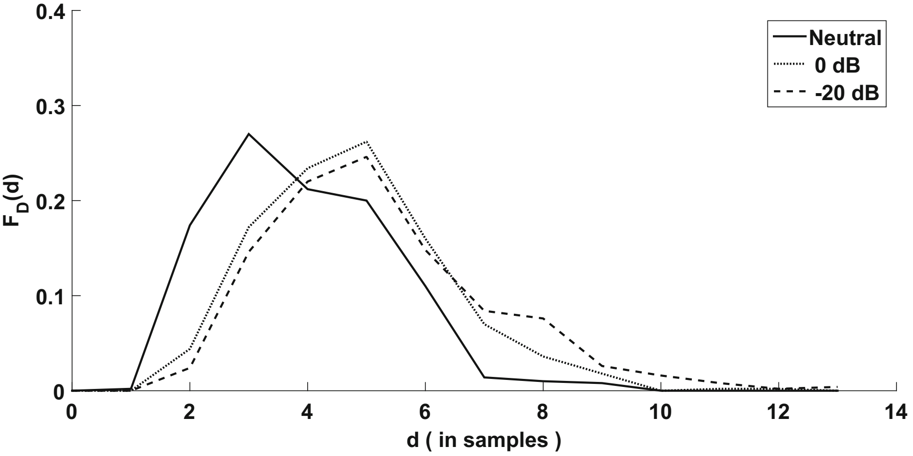


Fig. 3. Probability distribution of identification accuracies of speech signals of Neutral speech, Lombard speech at 0 dB and Lombard speech at −20 dB

The impact of LE on epoch extraction is demonstrated by comparing the identification accuracies [18,19] of neutral speech and speech under LE at 0 dB and −20 dB noise levels. Figure 2 shows the plot for EGG signals and Fig. 3 for speech signal, where 'd' is the identification accuracy and $F_D(d)$ is the probability distribution of the random variable 'D', representing the identification accuracies. For EGG as well speech signals, the distribution of neutral speech is found to be different from the speech under LE, where as the distributions of speech under LE at 0 dB and −20 dB noise levels are observed to have similar characteristics.

6 Summary and Conclusion

In the present work, a large database compared to the SUSAS dataset was developed to analyse the speech parameters under LE at different noise levels. The speech under LE was recorded in a simulated noisy condition and the neutral speech was recorded in a noise free environment. The speciality of the proposed dataset is that it contains EGG signal corresponding to each speech utterance in the dataset. To analyse the level of LE in the proposed dataset, a comparative study of the source parameters pitch and SoE, for speech under LE at different levels of noise was performed. The obtained trend of high pitch and low SoE for speech under LE for the developed dataset, is found to be consistent with the trend followed in literature [1]. In addition to this, the performance of epoch location extraction algorithm is compared for neutral speech and speech under LE. Based on the comparative study, the epoch extraction for speech under LE is found to have more deviation from the original epoch location, than that for neutral speech. The level of LE on the developed dataset is also confirmed using an MFCC-GMM based speaker recognition system. For the proposed as well as the standard datasets, the speaker recognition rates were high when the speaker recognition system was tested with neutral speech but was found to degrade when tested with speech under LE. Since the features of speech under LE are not reflected in the MFCC features, there should be new methods to compensate these features for speech under LE. Also, a new epoch extraction algorithm considering the features and characteristics of speech under LE should be developed to accurately measure the excitation source features. These modifications can lead to the development of a robust speaker recognition system for speech under LE.

References

1. Bapineedu, G., Avinash, B., Gangashetty, S.V., Yegnanarayana, B.: Analysis of lombard speech using excitation source information. In: Interspeech, pp. 1091–1094. Citeseer (2009)
2. Mahadeva Prasanna, S.R., Govind, D.: Analysis of excitation source information in emotional speech. In: INTERSPEECH, pp. 781–784 (2010)
3. Raja, G.S., Dandapat, S.: Speaker recognition under stressed condition. Int. J. Speech Technol. **13**(3), 141–161 (2010)

4. Hansen, J.H.L.: Analysis and compensation of speech under stress and noise for environmental robustness in speech recognition. Speech Commun. **20**(1–2), 151–173 (1996)
5. Furui, S.: 50 years of progress in speech and speaker recognition. In: SPECOM 2005, Patras, pp. 1–9 (2005)
6. Bapineedu, G.: Analysis of Lombard effect speech and its application in speaker verification for imposter detection. Ph.D. thesis, International Institute of Information Technology Hyderabad, India (2010)
7. Hagiwara, R.: Monthly mystery spectrogram. Linguistics Department, University of Manitoba, Canada (2006)
8. Ikeno, A., Varadarajan, V., Patil, S., Hansen, J.H.L.: Ut-scope: speech under Lombard effect and cognitive stress. In: Aerospace Conference, 2007 IEEE, pp. 1–7. IEEE (2007)
9. Hansen, J.H.L., Bou-Ghazale, S.E., Sarikaya, R., Pellom, B.: Getting started with SUSAS: a speech under simulated and actual stress database. In: Eurospeech, vol. 97, pp. 1743–1746 (1997)
10. Bořil, H., Pollák, P.: Design and collection of Czech Lombard speech database. In: Proceedings of Interspeech, vol. 5, pp. 1577–1580. Citeseer (2005)
11. Pravena, D., Govind, D.: Development of simulated emotion speech database for excitation source analysis. Int. J. Speech Technol. **20**, 327–338 (2017)
12. Shukla, S., Prasanna, S.R.M., Dandapat, S.: Stressed speech processing: human vs automatic in non-professional speakers scenario. In: 2011 National Conference on Communications (NCC), pp. 1–5. IEEE (2011)
13. Reynolds, D.A.: Speaker identification and verification using Gaussian mixture speaker models. Speech Commun. **17**(1), 91–108 (1995)
14. Pravena, D., Nandhakumar, S., Govind, D.: Significance of natural elicitation in developing simulated full blown speech emotion databases. In: 2016 IEEE Students on Technology Symposium (TechSym), pp. 261–265. IEEE (2016)
15. Govind, D., Mahadeva Prasanna, S.R., Pati, D.: Epoch extraction in high pass filtered speech using Hilbert envelope. In: INTERSPEECH, pp. 1977–1980 (2011)
16. Deepak, K.T., Prasanna, S.R.M.: Epoch extraction using zero band filtering from speech signal. Circ. Syst. Sig. Process. **34**(7), 2309–2333 (2015)
17. Ramesh, K., Mahadeva Prasanna, S.R., Govind, D.: Detection of glottal opening instants using Hilbert envelope. In: Interspeech, pp. 44–48 (2013)
18. Govind, D., Hisham, P.M., Pravena, D.: Effectiveness of polarity detection for improved epoch extraction from speech. In: 2016 Twenty Second National Conference on Communication (NCC), pp. 1–6. IEEE (2016)
19. Govind, D., Joy, T.T.: Improving the flexibility of dynamic prosody modification using instants of significant excitation. Circ. Syst. Signal Process. **35**(7), 2518–2543 (2016)

Classification of Alzheimer and MCI Phenotypes on MRI Data Using SVM

K.R. Kruthika[1](✉), Rajeswari[1], Akshay Pai[2], H.D. Maheshappa[1], and Alzheimer's Disease Neuroimaging Initiative

[1] Department of Electronics and Communication Engineering,
Acharya Institute of Technology, Bangalore, India
kr.kruthika@gmail.com, hdmappa@gmail.com

[2] Department of Computer Science, University of Copenhagen,
Copenhagen, Denmark

Abstract. Alzheimer disease (AD) is a common form of dementia affecting people older than the age of 65. Moreover, AD is commonly diagnosed by behavioural paradormants, cognitive tests, and is followed by brain scans. Computer Aided Diagnosis (CAD), applies medical imaging and machine learning algorithms, to aid in the early diagnosis of Alzheimer's severity and advancement from prodromal stages i.e. Mild Cognitive Impairment (MCI) to diagnosed Alzheimer's disease. In this work, SVM (support vector machine) is used for dementia stage classification. Anatomical structures of the brain were obtained from FreeSurfer's processing of structural Magnetic Resonance Imaging (MRI) data and is utilized for as features for SVM. To be more precise, the system is processed using T1-weighted brain MRI datasets consisting of: 150 mild cognitive impairment (MCI) patients, 80 AD patients and 130 normal controls (NC) obtained from Alzheimer Disease Neuroimaging Initiative (ADNI) database. The volumes of brain structures (hippocampus, medial temporal lobe, whole brain, ventricular, cortical grey matter, entorhinal cortex and fusiform) are employed as biomarkers for multi-class classification of AD, MCI, and NC.

Keywords: Alzheimer disease · Mild cognitive impairment · Normal control · Structural magnetic resonance imaging · FreeSurfer · Machine learning · SVM

Alzheimer's Disease Neuroimaging Initiative-Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf.

S.M. Thampi et al. (eds.), *Advances in Signal Processing and Intelligent Recognition Systems*, Advances in Intelligent Systems and Computing 678,
DOI 10.1007/978-3-319-67934-1_23

1 Introduction

Alzheimer's disease (AD) is a common form of dementia affecting millions of elderly people above the age of 65 worldwide. Before AD, ailments such as (MCI) serves as an intermediary phase between normal cognitive controls (NC) and AD. Furthermore, this MCI phase has a high conversion rate to AD. As a result, there is a need for the development of a sensitive, precise, and specific atrophy biomarkers for early detection of AD progression [1]. These new methods are needed to help researchers develop new treatments for Alzheimer's as discussed by Hua et al. [2]. Methods for early detection may further differ by the type of imaging biomarkers that can be applied [3–5]. For example, neuroimaging methods such as positron emission tomography (PET), functional magnetic resonance imaging (fMRI) and structural magnetic resonance imaging (MRI) are useful in evaluation of anatomical degradation caused by the disease [6–8]. Overtime, structural MRI of the brain has progressively become more employed in identifying structural changes in common aging diseases like Alzheimer's [9]. Structural brain MRI methods have the ability to utilize biomarkers that are presented in the image. These biomarkers are able to illustrate the structural differences for a healthy and diseased individuals. It is important to note that the methods may vary depending on the nature of the employed imaging biomarkers [10]. In spite of this, due to the ease of availability, non-persistent nature, and a high quality of MR images, they are the most suitable for differentiating changes in the brain anatomy due to disease development and progression.

As a result of the ubiquitous use of MRI in research and medicine, simultaneous advances in neuro-informatics have led to the materialization of many free and commercial image analysis software packages for the last 15 years. This includes but is not restricted to SPM, FSL, FreeSurfer, BrainVisa, Minboggle, NeuroQuant and NeurQlab. Premature diagnosis of AD by structural MRI studies is a challenging task because of its difficulty in quantifying patterns seen in the structural changes during early phases of AD or clinically normal phases [11]. Patients at the early stages of AD are classified as MCI, but not all MCI patients convert to AD. An analysis of research and clinical reports show that 5–10% of MCI patients convert to AD per year [12]. Voxel based (VBM) morphometry from high-resolution T1-weighted brain MRI data has been employed for diagnosis. Furthermore imaging biomarkers were obtained from the processed images such as grey matter concentration maps which are registered to a reference location for facilitating voxel by voxel comparisons across subjects [13]. In this work, we focus on the volumetric measurements of various brain structures as they have an impact on dementia diagnosis. Specifically, MCI is known to be effected by volume loss of brain structures like the hippocampus, MTI, the entorhinal cortex, and the total volume of the brain and is therefore exploited for classification.

Kloppel et al. applied support vector machine (SVM) to classify grey matter segments in T1-weighted MR scans obtained from diagnosed AD patients and the NCs obtained from two centers with dissimilar scanning equipment in order to generalize across different medical centers [14]. Magnin et al. proposed a new classification method of whole-brain (1.5-T) MRI to discriminate AD patients from NC subjects based on SVMs [15]. Here brain is divided into five regions of interest by using a previously developed anatomically labelled template for the brain and created a mask to exclude voxels of the skull. Vemuri et al. [16] developed a tool for Alzheimer's diagnosis through structural classification of MRI using SVM. As the dimension of these brain structures is collinear, it is essential to know which of them is more likely linked to severity of illness; the amount of atrophy in the other explains further variation in overall symptom severity. The studies in this field typically evaluate the diagnostic accuracy of AD and MCI patients with healthy control subjects. This study proposes volumetric measurement of hippocampus, medial temporal lobe, ventricles, amygdala, whole brain volume, cortical grey matter and entorhinal cortex and fusiform structures used as MRI biomarkers to predict different forms of dementia including the AD and the MCI. The MRI database scan for the proposed work has been taken from the AD Neuroimaging Initiative (ADNI) [17]. FreeSurfer software is employed to obtain hippocampal, MTL, and whole brain volumes, as well as ventricles, amygdala and cortical grey matter by cortical and sub-cortical segmentation. Furthermore, the SVM classification from LibSVM package is utilized for multi-class classification of AD, MCI and NC.

The organization of this paper is as follows. In Sect. 2, possible volume biomarkers of AD are discussed. In the first part of Sect. 3, the dataset and the FreeSurfer tool are briefly presented, followed by explaining the inner workings of SVM for Alzheimer's classifications. The whole process of classification is given in Fig. 1. Section 4 is devoted for discussing the performance of the presented method. Finally, in Sect. 5 the conclusion and the future work are communicated.

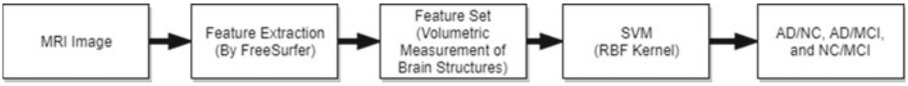

MRI Image
Feature Extraction (By FreeSurfer)
Feature Set (Volumetric Measurement of Brain Structures)
SVM (RBF Kernel)
AD/NC, AD/MCI, and NC/MCI

Fig. 1. Flowchart

2 Volume Biomarkers of AD

Manual volumetric measurements of brain structures is regarded as "the gold standard" for detecting symptoms of AD. However, it is time consuming and has an operator bias. In comparison, automatic measuring methods such as voxel-based morphometry (VBM) are fast and are extensively employed in the field [23–25]. However, this method is not to define every gyrus in the brain and is criticized by some to have confounding issues [26]. Lies et al. has addressed some of these issues where it is found that a VBM method is measuring the same effects as "the gold standard" concerning to the subcortical brain structures [27]. Overall, major structures in the brain like hippocampus, medical temporal lobe, ventricles, amygdala, cortical grey matter, entorhinal cortex and the whole brain volume are investigated for indications of atrophy that lead to AD.

2.1 Hippocampus

The hippocampus creates the majority of the temporal lobe and is commonly used for AD diagnosis. Moreover, hippocampal atrophy is a well-known cause of dementia [9]. Specifically, hippocampal atrophy differentiates the three main disease stages of AD, MCI and NC [21]. It is also speculated that a low hippocampus volume can be utilized as a new diagnostic criterion for MCI patients with high risks of AD conversion [11].

2.2 Medial Temporal Lobe

The medial temporal lobe (MTL) region contains structures that are key in long-term memory. As a result, a structural MRI of the MTL's atrophy is an effective indicator for the initial diagnosis of AD. Visser et al. reported these results in 1999 among 45 patients in their study [17].

2.3 Ventricles

Ventricles are cavities in the cerebral hemispheres filled with cerebrospinal fluid. Furthermore, their volume variations indicate the existence of AD. These cavities are found to expand in size steadily in AD patients [20]. In particular, Apostolova et al. has reported that the use of cerebral ventricular volume for measurement of AD development. They claimed that the hemispheric atrophy rate calculated by ventricular enlargement correlates strongly with changes on cognitive tests and are able to capture significant variations among levels the stages of Alzheimer's [18].

2.4 Amygdala

Amygdala is a primary limbic structure anatomically interconnected with the neocortex. In particular, the amygdala serves as a structure for how emotions are processed. In cases of AD, neural lost and alterations in glial cell population have been reported. In support, Poulin et al. reported the magnitude of amygdala atrophy is considerable in AD stages [19].

2.5 Whole Brain

Volumetric MRI studies have found relationships between increasing age and decreasing brain volumes. In particular, there is an age-correlated decrease in hippocampal, temporal, frontal lobe structure volumes, and an increase in cerebrospinal spaces [20]. Moreover, there are more sensitive predictors of AD and MCI are achievable by exploiting the whole brain's atrophy rate along with the hippocampal volume [21].

2.6 Cortical Grey Matter

MRI measurements of cortical grey matter and abnormal white matter are independently connected with dementia severity. Both biomarkers have their own contributions to the performance in MCI domains as well. For example, quantitative MRI provides a strong conformation that cortical grey matter volume are related to atrophy and abnormal white matter volume are separately related to the dementia severity in AD subjects [22].

2.7 Entorhinal Cortex

Entorhinal cortex is a key pre-processor that stimulates the nearby hippocampus. It serves as an area for memory and navigation. Examinations have confirmed this assumption; also, few observations illustrate that entorhinal cortex is the primary part which is affected in MCI cases even earlier than hippocampus.

3 Materials and Methods

3.1 Data

Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). The ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer's disease (AD). The ADNI was collectively launched by six non-profit organizations in 2003: the National Institute on Aging (NIA), the National Institute of Biomedical Imaging and Bioengineering (NIBIB), the Food and Drug Administration (FDA), private pharmaceutical companies and available at adni.loni.usc.edu. It aims to assess whether structural MRI, positron emission tomography (PET), biomarkers, as well as clinical and neuropsychological assessments can be collectively measure the progression of MCI and early AD. The dataset is divided into categories of AD, MCI and NC, where MCI consists of EMCI and LMCI as shown in Table 1.

Table 1. Overview of the MRI dataset

Class	# of subjects	# of males/females	Age (mean ± std)
AD	200	103/97	75.40 ± 7.61
EMCI	150	77/73	73.24 ± 6.19
LMCI	150	73/77	74.10 ± 7.73
NC	200	73/102	76.49 ± 6.78

3.2 FreeSurfer Processing

FreeSurfer is one of the most widely used software today for volumetric analysis of the brain. It is indeed a set of tools for cortical analysis and visualization and sub-cortical segmentation of MRI data [28]. Accurate and reliable segmentation is a necessity for volumetric analysis of dementia disease. Here, sub-cortical and cortical volumetric measurements were computed by FreeSurfer (version 5.3.0) using atlas based labelling of region of interest (ROI) [29]. Statistical output files generated during FreeSurfer processing stream was used to obtain hippocampus volume and intra-cranial volume (ICV). The volumes of medial temporal lobe, ventricles, amygdala, CGM, entorhinal cortex, fusiform, and the whole brain was computed using anatomical ROI segmentation analysis of their given file: aparc.a2009s+aseg.mgz. The volume of each structure is found by counting the voxels of each of these coloured and labelled structure using an .mgz image that FreeSurfer outputs by using MATLAB. Each volume calculated was then normalized by dividing them with the intra-cranial volume (ICV). The ICV was found from surfer.nmr.mgh.harvard.edu with three aseg.stat files available at 7 head-sized corrections to reduce inter-individual variation. FreeSurfer processing is computationally expensive and takes several hours to process a single image. Therefore, in order to reduce computational time, eight images are processed in parallel using GNU Parallel on an 8 core machine.

Support Vector Machine. Support vector machine is a machine learning method that classifies binary classes by finding a class boundary. This boundary, the hyper plane, is used to find the maximum margin in the given training data. The training data samples along the hyper planes near the class boundary are called support vectors and the margin is the distance between the support vectors and the class boundary hyperplanes. The SVM classifier is based on the concept of decision planes that define decision boundaries. A decision plane is one that separates between assets of objects having different class memberships. Furthermore, a classification task usually involves training and testing data, which consists of some data instances. Where each instance in the training set contains one target value (class labels) and several attributes (features). SVMs have an advantage that its objective function is convex; however, it can only guarantee to converge to a local minimum. Moreover, it is fundamentally a two-class classifier. One commonly used approach to tackle problems involving more than two classes is the one-versus-the-rest approach and is as followed:

Given a training data set with labels $\{(x_1, y_1), ...(x_n, y_n)\}$ where $x_i \in R^n$ and $y_i \in \{+1, -1\}$ and a non-linear map $\phi()$, that maps to a higher dimensional space, R^n R^H the SVM technique solves:

$$\min_{\omega,\xi_i,b} \{\frac{1}{2}\|\omega\|^2 + C \sum \xi_i\} \tag{1}$$

Subject to the constraints:

$$y_i(\phi^T(x_i)w + b) \geq 1 - \xi_i, i = 1, 2...n \tag{2}$$

$$\xi_i \geq 0, i = 1, 2...n \tag{3}$$

specifically w and b define linear classifiers in a feature space. According to Cover's theorem, a non-linear mapping function ϕ is performed allowing transformed samples to be more likely linearly separable [30]. A regularizer parameter C allows control over penalty assignment to errors model. Slack variable ξ are introduced to account for non-separable data involved with permitted errors

Owing to the higher dimensionality of vector variable w, the primal function in Eq. (1) is solved by its Lagrangian dual problem which consists of maximizing:

$$L_d = \sum_i \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j \phi(x_i) \phi(x_j) \tag{4}$$

subject to constraints

$$\sum_i \alpha_i y_i = 0, i = 1, 2...n \tag{5}$$

$$C \geq \alpha_i \geq 0, i = 1, 2...n \tag{6}$$

where α_i are Lagrange multipliers corresponding to Eq. (2). It can be noted that all ϕ mappings used in the SVM learning occur in the form of inner products. Furthermore, Boster et al. proposed a way to model more complicated relationships by replacing the inner product with a kernel function (such as a Gaussian radial basis function, polynomial kernel or a linear kernel) [31]. This allows us to define a kernel function K where the inner products in the original space (x_i, x_j) replaced with inner products in the transformed space $[\phi(x_i).\phi(x_j)]$:

$$K(x_i, x_j) = \phi(x_i).\phi(x_j) \tag{7}$$

This kind of kernel function allows us to simplify the solution of the dual problem considerably. This is because it avoids the computation of the inner products in the transformed space $[\phi(x_i).\phi(x_j)]$. Though ϕ mapping can be explicitly expressed for a linear or polynomial kernel, there is no explicit form of ϕ mapping corresponding to the Gaussian kernel. Moreover, it can be demonstrated that the expansion is an infinite-dimensional functional [32]. Mercer's theorem avoids to explicitly calculate ϕ in these cases, and then, by introducing (7) into (4), the dual problem can be finally stated as [33]:

$$L_d = \sum_i \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j K(x_i, x_j) \tag{8}$$

After the dual problem is solved, $w = \sum_{i=1}^{n} \alpha_i y_i \phi(x_i)$ and express the final result as a decision $f(x)$. Where any test data x is in the original (lower) dimensional feature space:

$$f(x) = sgn((\sum_{i=1}^{n} \alpha_i y_i K(x_i, x_j) + b)) \tag{9}$$

Furthermore, b can be easily computed from the α_i that are neither zero nor C.

The shape of the discriminant function depends on the kind of kernel functions adopted. A common kernel type that fulfills Mercer's condition is the Gaussian radial basis function where γ controls the shape of the peaks and the data points are transformed to a higher dimension:

$$K(x_i, x_j) = exp(-\gamma \|x_i - x_j\|^2), \gamma > 0 \tag{10}$$

where γ is a free parameter inversely proportional to the width of the Gaussian kernel.

A small γ means a Gaussian with a large variance resulting a stronger influence of x_j. In other words, if x_j is a support vector, a small γ implies the class of this support vector will have influence that has a high bias on deciding the class of the vector x_i even if the distance between them is large. If γ is large, then variance is small implying that the support vector does not have a wide-spread influence (a low bias). A low bias is utilized because the cost of misclassification is penalized heavily. However, a large γ leads to a high bias and low variance models and vice versa.

The FreeSurfer tool is used to take volume of different brain regions such as medial temporal lobe, ventricles, amygdala, cortical grey matter (CGM), entorhinal cortex, and fusiform in each subject. In the training data, each row is a sample, and the columns consists the above stated feature and labels for each sample. For example, hippocampus training data for AD vs. NC classification consists 400 rows and each row represents a sample/subject; one column consists the feature for each sample; and one more column with labels: here +1 for AD and −1 for NC. All training data is prepared in a similar manner for all the aforementioned brain regions.

The data is scaled before SVM is applied [34], The main advantage of scaling is to avoid attributes in greater numeric ranges dominating those in smaller numeric ranges. Another advantage is to avoid numerical difficulties during the calculation. In order to develop an SVM, penalization parameter C; and kernel parameter γ must be tuned. The best C and γ hyper-parameters are found using Grid-Search. Grid search is when given a set of models (which differ from each other in their parameter values, which lie on a grid), train each of the models and evaluate it using 5 - fold cross-validation. Then select the one that performed best. The best C value is 512 and γ is 0.03125. Finally, from 700 subjects' data, 75% training and 25% testing data are taken randomly, and used for training and then to evaluate the model's performance respectively. Here we have implemented SVM using the libSVM [35] software package.

4 Results and Discussions

The simulated results presented are obtained using an 8 Core machine with 8 Giga Bytes of random access memory and 3 Mega Bytes of cache.

The area under the curve (AUC) of a two-class classification of combinations of the prodromal stages of dementia are shown in Table 2. AUC analysis, a commonly chosen metric, is chosen to compare the performance of classification models. The predominate reason for using AUC as an alternative to accuracy is that it is not as sensitive to differences between the class distribution within the training and test samples [36,37]. To be precise, an AUC driven analysis helps in deciding a correct model when one may have been trained on a skewed data set.

Table 2. AUC of different combinations of the stages of dementia using SVM

Brain structure	AD/ MCI	AD/ NC	MCI/ NC	LMCI/ AD	EMCI/ AD	EMCI/ LMCI	EMCI/ NC	LMCI/ NC
Hippocampus	0.7913	0.9575	0.6409	0.5294	0.8794	0.8114	0.3694	0.7184
Medial temporal lobe	0.7787	0.915	0.5939	0.6658	0.8559	0.6523	0.5069	0.6376
Ventricles	0.6232	0.6569	0.5225	0.5481	0.6971	0.5068	0.3944	0.5543
Amygdala	0.7899	0.8382	0.5331	0.6604	0.8647	0.5977	0.4486	0.6212
Cortical grey Matter	0.8123	0.8333	0.4722	0.7166	0.8	0.5909	0.4833	0.6111
Whole Brain	0.7831	0.8448	0.5498	0.7594	0.8882	0.6182	0.6153	0.596
Entorhinal cortex	0.6541	0.7059	0.5397	0.5856	0.8118	0.6682	0.4208	0.5808
Fusiform	0.7451	0.799	0.5311	0.6123	0.7912	0.6795	0.4264	0.7285
Combined	0.6457	0.7974	0.6157	0.5642	0.7441	0.6886	0.5986	0.6717

From Table 2, features from the hippocampus are shown to act as better discriminators for most stages of dementia except for LMCI/AD and EMCI/NC. This is in support of the argument that, the hippocampus acts as a sensitive biomarker for earlier stages of dementia. The second highest performing biomarker is utilizing the medial temporal lobe (MTL). Though MTL as a biomarker does not perform as the best discriminator for any individual combination, it performs the best on average. Ventricles and entorhinal cortex structures are shown to be below average discriminators, as they do not even discriminate one combination of dementia stage. Moreover, despite combining all the biomarkers, it does not perform as the best discriminators overall and only excels at EMCI/NC classification. The CGM biomarker performs well for AD/MCI and LMCI/AD. The whole brain performs well for AD/LMCI and EMCI/AD, and EMCI/NC. The Fusiform performs best for LMCI from NC. The combined features perform well for MCI, EMCI discrimination from NC. The performance curve for AD/MCI, AD/NC and MCI/NC using the hippocampus features are shown in Figs. 2, 3, 4, 5, 6 and 7.

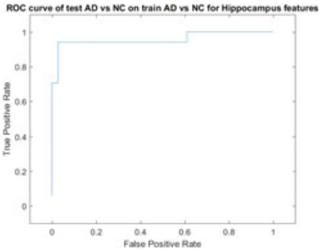


Fig. 2. ROC curve plotted for Hippocampus features of AD vs. NC

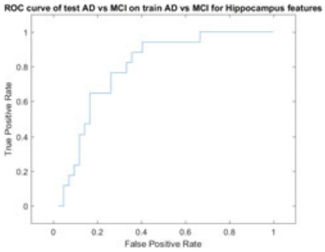


Fig. 3. ROC curve plotted for Hippocampus features of AD vs. MCI

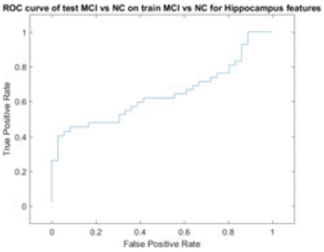


Fig. 4. ROC curve plotted for Hippocampus features of MCI vs. NC

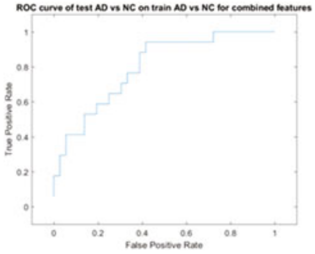


Fig. 5. ROC curve plotted for Combined features of AD vs. NC

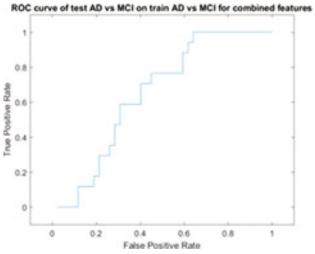


Fig. 6. ROC curves plotted for Combined features of AD vs. MCI

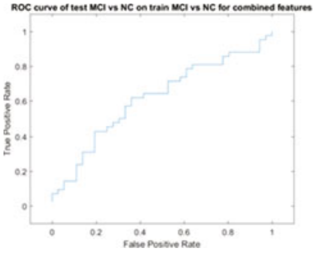


Fig. 7. ROC curve plotted for Combined features of MCI vs. NC

5 Conclusions and Future Work

In this study we examined the accuracy and reliability of multi class classification based on ROC using volumetric measurements of different brain structures for an accurate diagnosis of dementia stages. Hippocampal volume measurements are the best discriminate for transitions of: AD from NC, AD from MCI, and NC from MCI. The results obtained are satisfactory and are based on a database of hippocampus features. This database consisted of: 400 images for AD vs. NC, 500 images for AD vs. MCI, and 500 images for NC vs. MCI. Moreover, we were able to achieve an AUC value 95.75%, 79.13% and 64.09% respectively.

For future work, will be on the use of raw data to classify stages of dementia using a deep learning approach such as a convolutional neural network. Furthermore, we would like to explore the performance of utilizing combined features of hippocampus, CGM, and volume of the entire brain and how they complement each other on the several stages of dementia classification.

Acknowledgements. Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd. and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

References

1. Liu, Y., Paajanen, T., Zhang, Y., et al.: Combination analysis of neuropsychological tests and structural MRI measures in differentiating AD, MCI and control groups the AddNeuroMed study. Neurobiol. Aging **32**(7), 1198–1206 (2011)
2. Hua, X., Leow, A., Lee, S., et al.: 3D characterization of brain atrophy in Alzheimer's disease and mild cognitive impairment using tensor-based morphometry. NeuroImage **41**, 19–34 (2008)
3. Hua, X., Lee, S., Yanovsky, I., et al.: Optimizing power to track brain degeneration in Alzheimer's disease and mild cognitive impairment with tensor-based morphometry: an ADNI study of 515 subjects. NeuroImage **48**, 668–681 (2009)
4. Markiewicz, P., Matthews, J., Declerck, J., et al.: Robustness of multivariate image analysis assessed by resampling techniques and applied to FDG-PET scans of patients with Alzheimer's disease. NeuroImage **46**, 472–485 (2009)
5. Walhovd, K., Fjell, A., Amlien, I., et al.: Multimodal imaging in mild cognitive impairment: metabolism, morphometry and diffusion of the temporalparietal memory network. NeuroImage **45**, 215–223 (2009)
6. Tripoliti, E.E., Fotiadis, D.I., Argyropoulou, M., et al.: A six stage approach for the diagnosis of the Alzheimers disease based on fMRI data. J. Biomed. Inform. **43**(2), 307–320 (2010)

7. Shin, J., Lee, S.-Y., Kim, S.J., et al.: Voxel-based analysis of Alzheimer's disease PET imaging using a triplet of radiotracers: PIB, FDDNP, and FDG. NeuroImage **52**, 488–496 (2010)
8. Frisoni, G.B., Fox, N.C., Jack, C.R., et al.: The clinical use of structural MRI in Alzheimer disease. Nat. Rev. Neurol. **6**, 67–77 (2010)
9. He, Y., Evans, A.: Magnetic resonance imaging of healthy and diseased brain networks. Front. Hum. Neurosci. **2014**(8), 890 (2015). doi:10.3389/fnhum.2014.00890
10. Johnson, K.A., Fox, N.C., Sperling, R.A., et al.: Brain imaging in Alzheimer Disease. Cold Spring Harbor Perspect. Med. **2**, a006213–a006213 (2012)
11. McGeown, W.J., Shanks, M.F., Forbes-McKay, K.E., et al.: Patterns of brain activity during a semantic task differentiate normal aging from early Alzheimer's disease. Psychiatry Res. Neuroimag. **173**, 218–227 (2009)
12. Torpy, J.M.: Mild cognitive impairment. JAMA **302**, 452 (2009)
13. Schmitter, D., Roche, A., Marchal, B., et al.: An evaluation of volume-based morphometry for prediction of mild cognitive impairment and Alzheimer's disease. NeuroImage Clin. **7**, 7–17 (2015)
14. Kloppel, S., Stonnington, C.M., Chu, C., et al.: Automatic classification of MR scans in Alzheimer's disease. Brain **131**, 681–689 (2008)
15. Magnin, B., Mesrob, L., Kinkingnhun, S., et al.: Support vector machine-based classification of Alzheimers disease from whole-brain anatomical MRI. Neuroradiol. **51**, 73–83 (2009). 17 24 P
16. Vemuri, P., Gunter, J.L., Senjem, M.L., et al.: Alzheimer's disease diagnosis in individual subjects using structural MR images: validation studies. NeuroImage **39**, 1186–1197 (2008)
17. Visser, P.J., Scheltens, P., Verhey, F.R.J., et al.: Medial temporal lobe atrophy and memory dysfunction as predictors for dementia in subjects with mild cognitive impairment. J. Neurol. **246**, 477–485 (1999)
18. Nestor, S.M., Rupsingh, R., Borrie, M., et al.: Ventricular enlargement as a possible measure of Alzheimer's disease progression validated using the Alzheimer's disease neuroimaging initiative database. Brain **131**(9), 2443–2454 (2008)
19. Poulin, S.P., Dautoff, R., Morris, J.C., et al.: Amygdala atrophy is prominent in early Alzheimer's disease and relates to symptom severity. Psychiatry Res. Neuroimag. **194**(1), 7–13 (2011)
20. Sluimer, J.D., van der Flier, W.M., Karas, G.B., et al.: Whole-brain atrophy rate and cognitive decline: longitudinal MR study of memory clinic patients 1. Radiology **248**(2), 590–598 (2008)
21. Henneman, W.J.P., Sluimer, J.D., Barnes, J., et al.: Hippocampal atrophy rates in Alzheimer disease added value over whole brain volume measures. Neurology **72**(11), 999–1007 (2009)
22. Stout, J.C., Jernigan, T.L., Archibald, S.L., et al.: Association of dementia severity with cortical gray matter and abnormal white matter volumes in dementia of the Alzheimer type. Arch. Neurol. **53**(8), 742–749 (1996)
23. Yokum, S., Stice, E.: Initial body fat gain is related to brain volume changes in adolescents: a repeated-measures voxel-based morphometry study. Obesity **25**(2), 401–407 (2017)
24. Riddle, K., Cascio, C.J., Woodward, N.D.: Brain structure in autism: a voxel-based morphometry analysis of the Autism Brain Imaging Database Exchange (ABIDE). Brain Imaging Behav. **11**(2), 541–551 (2017)
25. Chen, Q., et al.: Brain gray matter atrophy after spinal cord injury: a voxel-based morphometry study. Front. Hum. Neurosci. **11**, 211 (2017)

26. Focke, N.K., Trost, S., Paulus, W., Falkai, P., Gruber, O.: Do manual and voxel-based morphometry measure the same? a proof of concept study. Front. Psychiatry **5**, 39 (2014). doi:10.3389/fpsyt.2014.00039
27. Clerx, L., et al.: Can FreeSurfer compete with manual volumetric measurements in Alzheimer's Disease? Curr. Alzheimer Res. **12**(4), 358–367 (2015)
28. Fischl, B.: Freesurfer. NeuroImage **62**, 774–781 (2012)
29. Fischl, B., Salat, D.H., Busa, E., et al.: Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. Neuron **33**(3), 341–355 (2002)
30. Cover, T.M.: Geometrical and statistical properties of systems of linear inequalities with application in pattern recognition. IEEE Trans. Electron. Comp. **14**, 326–334 (1965). (reprinted. In: Mehra, P., Wah, B. (eds.) Artificial Neural Networks: Concepts and Theory. IEEE Computer Society Press, Los Alamitos, California (1992))
31. Boser, B.E., Guyon, I.M., Vapnik, V.N.: A training algorithm for optimal margin classifiers. In: Proceedings of the Fifth Annual Workshop on Computational Learning Theory, Pittsburgh, Pennsylvania, USA, 27–29 July 1992, pp. 144–152 (1992)
32. Schölkopf, B., Smola, A.: Learning with Kernels-Support Vector Machines, Regularisation, Optimization and Beyond. The MIT Press Series, Cambridge (2001)
33. Mercer, J.: Functions of positive and negative type and their connection with the theory of integral equations. Philos. Trans. Roy. Soc. A **209**(441–458), 415–446 (1909)
34. Sarle, W.S.: Neural network FAQ (1997). ftp://ftp.sas.com/pub/neural/FAQ.html. Periodic posting to the Usenet newsgroup comp.ai.neural-nets
35. Chang, C.-C., Lin, C.-J.: LIBSVM: a library for support vector machines. ACM Trans. Intell. Syst. Technol. **2**, 27:1–27:27 (2011)
36. Hanley, J.A., McNeil, B.J.: The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology **143**, 29–36 (1982)
37. Rathore, S., Habes, M.: A review on neuroimaging-based classification studies and associated feature extraction methods for Alzheimer's disease and its prodromal stages. NeuroImage (2017). doi:10.1016/j.neuroimage.2017.03.057

Real-Time Traffic Light Signal Recognition System for a Self-driving Car

Nakul Agarwal[1(✉)], Abhishek Sharma[2], and Jieh Ren Chang[3]

[1] Undergraduate, Computer Science and Engineering,
The LNM Institute of Information Technology, Jaipur, India
15ucs078@lnmiit.ac.in

[2] Department of Electronics and Communication Engineering,
The LNM Institute of Information Technology, Jaipur, India
abhisheksharma@lnmiit.ac.in

[3] Department of Electronic Engineering, National Ilan University, Yilan, Taiwan
jrchang@niu.edu.tw

Abstract. In this paper, the implementation of image recognition for traffic light signal recognition system is demonstrated. The detection of traffic light signal is an essential step for a self-driving car. Here we present a method for the recognition of traffic lights using image processing and controlling the vehicle accordingly. The algorithm developed in this research work is tested and processed using a Raspberry Pi board. The input-output modules such as camera, motors and chassis of the model car are all integrated together so they can perform as a single unit. For processing the image on real-time, OpenCV is used as an API to perform essential steps in the detection of signal like capturing, resizing, thresholding and morphological operations. Contour detection on a binary image has further been used for object detection. The algorithm has been tested with Valgrind profiling tools Callgrind and Cachegrind.

Keyword: Self-driving autonomous car · Color detection · Image processing · Opencv · Binary image · Contour detection · Convex hull · Raspberry Pi

1 Introduction

Self-driving cars are the new buzz for all the big automobile companies. With big research groups like from Google, Uber, Tesla, etc. attached, there have been significant advances towards this dream. The biggest problem these autonomous cars face is recognizing driving patterns and traffic signals. In [1], various problems faced by vision based cars have been discussed. Detection of signals and signs is faced with the issue of obstruction and lighting. [2] shows that even a red balloon on the side of the road or the setting sun has been confused with the red-light signal. There are cases when traffic signs get obstructed because of trees. In adverse weather conditions, the feature of object detection faces a major setback as discussed in [3]. These cars also need to guess the driving style and patterns of human drivers. They need to understand the patterns of driving and how driving behavior could be different in places like near traffic lights. This is one of the many other obvious reasons which makes traffic light signal recognition

S.M. Thampi et al. (eds.), *Advances in Signal Processing and Intelligent Recognition Systems*, Advances in Intelligent Systems and Computing 678,
DOI 10.1007/978-3-319-67934-1_24

one of the most essential feature for an autonomous car. The problem of recognizing traffic light has been dealt differently with different levels of outcomes. In [1] various vision based attempts for traffic light recognition have been discussed. The problem of different lighting outside at different times of the day is discussed in [4]. In [5], movement patterns of other vehicles have been studied to guess the current state of traffic light. In [6], adaptive background suppression filters have been applied to solve the problem. Deep learning using convolutional neural networks has been used in [7] to detect traffic light signals. In [8], machine learning classifiers like LDA, kNN and SVM have been implemented for dealing with this problem. Here, an algorithm to recognize traffic light signals using image processing has been developed. Contour detection surrounding any continuous high intensity pixel values based on the algorithm at [9] have been used. In this case, binary images are used, where high intensity means pixel value ‘1’ and low intensity means pixel value ‘0’. This binary image has been produced using thresholding on three (r, g, b) channels. Range of accepted (r, g, b) values for each signal has been predefined on the basis of environment. Further preprocessing has been performed before performing contour detection. Morphological operations [10–14] which are based on set theory have been performed to remove disturbance and tuning the detected objects. Median filtering [14–19] has been performed to further smooth out detected objects. Detected contours hulls have been approximated into convex hulls.

2 The Method

The purpose of this algorithm is to process images in such a way that the presence of any traffic light signal is detected, a convex hull is drawn around the region in the image which has the signal present and the color detected is passed. According to this passed signal, the raspberry pi board will activate its pins, pass current to the motors which in turn rotates the wheels attached to the chassis. If color signal is “RED”, then motors on the chassis don’t get any current and they stop moving. If color signal passed is “GREEN”, then pins pass current with high duty cycle which makes the motors rotate. If “YELLOW” is the passed current signal, then the pins pass current of a lower duty cycle, which makes the motors rotate at a lower r.p.m.

The convex contours hulls are drawn on the image captured. For this purpose, the OpenCV library function findContour (based on the algorithm discussed in [9]) which accepts a binary image is used. A contour is drawn around group of ‘1’s or the white blocks in the binary image. To generate binary image for the corresponding color, inRange function from the OpenCV library is called. The binary image generated by this function may not be accurate to the requirement of contour detection and may contain many other small disturbances. To get rid of these disturbances, various filters as discussed later have been used. The stepwise code flow diagram of the complete method is given in Fig. 1.

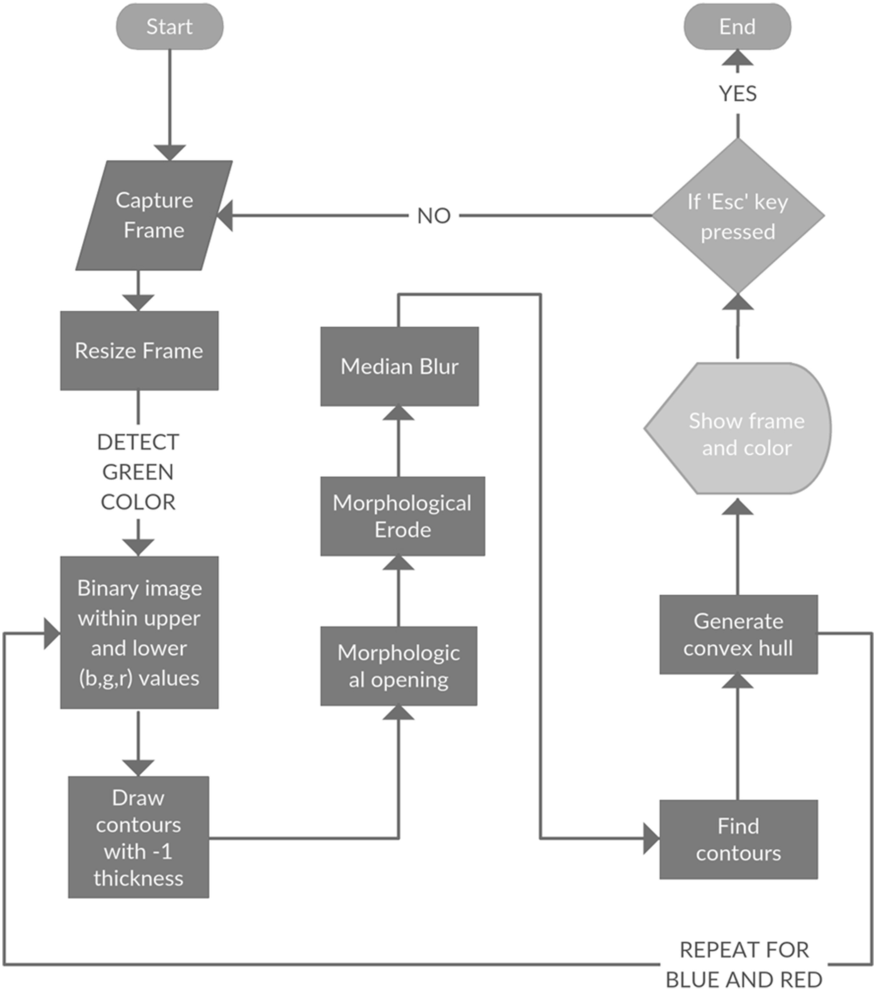
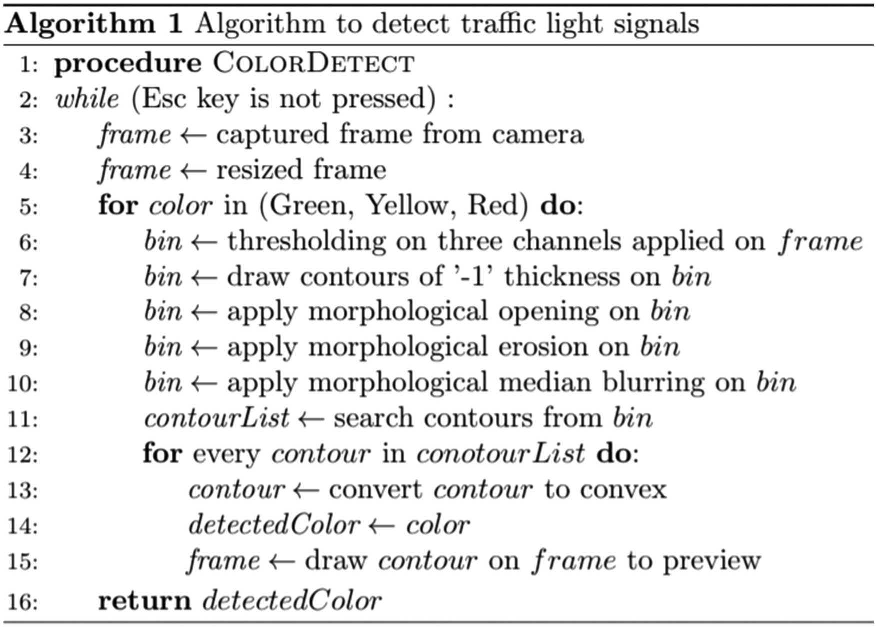


Fig. 1. Code-flow diagram for the algorithm. Diagram created using Creately Tool [27].

Pseudo-code for the detection of traffic light signals is shown in Fig. 2. Step wise description of each step of code-flow diagram in Fig. 1 and pseudo-code from Fig. 2 is given in subsequent parts of this section.

For this program, the codes are written in python language because of its readability, writability and vast open-source backing. OpenCV library from python is being used to perform image processing tasks. All images are handled as Numpy arrays which is a Python extension module as it provides faster and easier matrix handling and operations.



```
Algorithm 1 Algorithm to detect traffic light signals
procedure ColorDetect
while (Esc key is not pressed) :
    frame ← captured frame from camera
    frame ← resized frame
    for color in (Green, Yellow, Red) do:
        bin ← thresholding on three channels applied on frame
        bin ← draw contours of '-1' thickness on bin
        bin ← apply morphological opening on bin
        bin ← apply morphological erosion on bin
        bin ← apply morphological median blurring on bin
        contourList ← search contours from bin
        for every contour in conotourList do:
            contour ← convert contour to convex
            detectedColor ← color
            frame ← draw contour on frame to preview
    return detectedColor
```

Fig. 2. Pseudo-code for the algorithm

2.1 Hardware Setup

The whole experiment has been executed on a Raspberry Pi camera module has been installed to provide input to the Raspberry Pi board running the color detection algorithm.

On the basis of the detected traffic signal color, the program dictates Raspberry Pi to activate its General-purpose input/output pins (GPIO Pins). These GPIO pins are connected to RC motors attached to the car chassis which in turn rotate the wheels of chassis with an r.p.m. depending upon the duty cycle of the current passed. The whole setup is demonstrated in Fig. 3.

2.2 Capturing Frames

A USB camera with frame frequency 30FPS was used to capture frames. The OpenCV assists in capturing frames from the camera. A VideoCapture object is created which helps to capture live stream with a camera. Each frame is captured as a numpy array which provides faster array calculation. Every captured frame is resized to a smaller size with width of 300 pixels to reduce computational time. On each frame, image processing is applied to detect traffic light signal. Each frame is first checked for the presence of a green light signal, then for blue and then for red. This way, highest priority is given to red signal for safety.

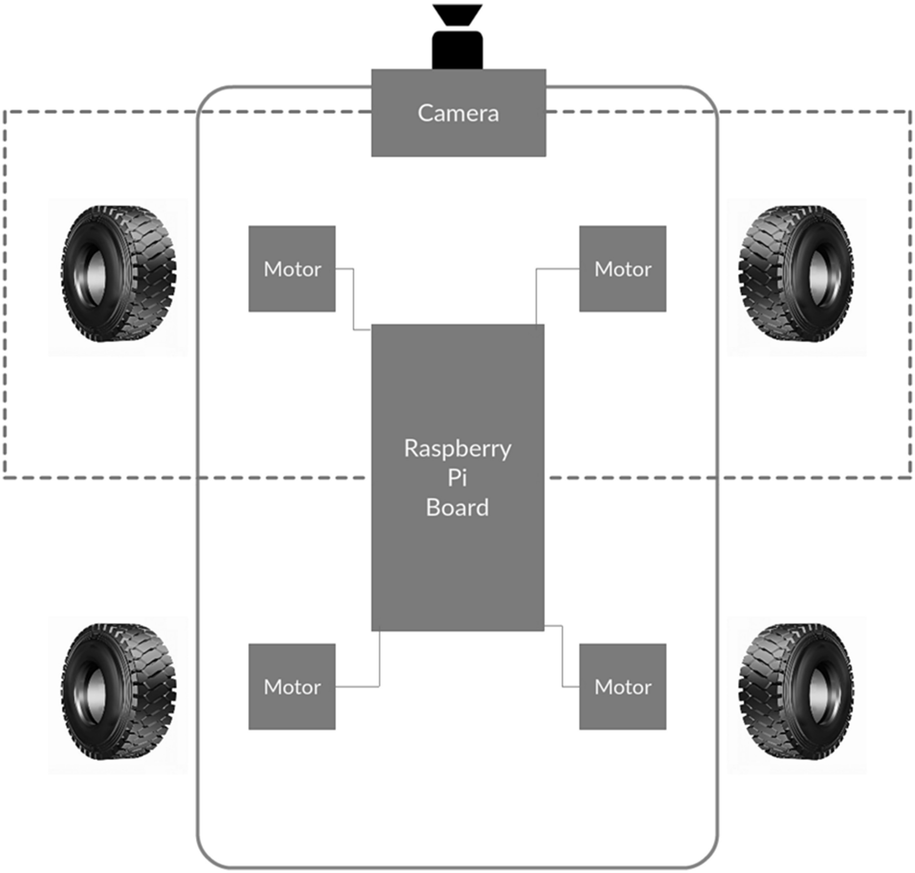


Fig. 3. Hardware setup. Diagram created using Creately Tool [27].

2.3 Generation of Binary Images

The captured RGB frames are needed to be converted into binary images based on a specific range of (r, g, b) values. These binary images should have pixel value 1 for pixels in image which lie in a specified (r, g, b) range and value 0 if they do not. The range of (r, g, b) values predefined according to the color to be detected and surrounding environments is present. To perform this thresholding on multiple channels, the inRange function from the OpenCV library is called.

2.4 Removal of Disturbances

The binary images obtained cannot be passed for contour detection yet. Some pixels may be present, which have their values lying in the predefined (r, g, b) range but are not needed to be detected. Contours detection is needed around only clusters or blocks of ‘1’ pixel values. To obtain such a binary image, some image processing operations must be applied on the numpy array.

There are cases when the clusters of ‘1’ valued pixels are not continuous, i.e. they may have some pixels having values ‘0’ inside them. To solve this problem, contours with thickness ‘−1’ of color white (pixel value ‘1’) can be drawn.

Now to remove small disturbances in this binary image, use of morphological operations [12–14] based on mathematical morphology [10, 11, 14] is preferred. Mathematical morphology unlike traditional image processing techniques, treats an image as an ensemble of sets. It needs two inputs, i.e. image and a structuring element or kernel which decides the nature of the operation. The primary morphological operations are dilation and erosion, which are based on operations like translation, set union and set intersection. Morphological opening function has been used to remove the small disturbances in this binary image. Opening function is another name for erosion function followed by dilation. For a kernel, we will use a 3 × 3 identity matrix. Also, there are some white blocks or clusters which remain connected to each other. This can be removed by using the erosion morphological operator. Same kernel defined earlier is being used.

Finally, a median blur [14–19] is applied to the binary image. A median blur goes through each pixel value and replaces it with the median of all the pixel values lying in the surrounding kernel of a specified size. A kernel of size 7 × 7 is used. Median blur is effectively useful for salt-and-pepper noise.

2.5 Convex Contour Hulls

Contours are useful for shape analysis and object detection from binary images. Contours are found from the generated binary image for all the continuous points or blocks and are stored in a list. If a contour is found, the detected color is changed. Every contour stored in the list is approximated to a convex hull in case the contours are concave.

Each contour from the list is then drawn onto the original frame and the frame is previewed.

3 Results

This experiment was run completely on a Raspberry Pi Model B. The algorithm worked satisfactorily in recognizing traffic light colors. Each frame is previewed with convex hull drawn on the detected signal. All the result is achieved with a time minimal time lag ~ 1 s.

The algorithm was further tested with Valgrind profilers tools i.e. Cachegrind and Callgrind. The Cachegrind profiler tells the total cache hits by the program. The output is presented in Table 1.

Table 1. Cachegrind profiling results

Cache Type	Program Total Type
Ir	2,456,080,245
I1mr	1,436,113
ILmr	92,988
Dr	1,039,494,031
D1mr	6,264,598
DLmr	411,331
Dw	343,926,705
D1mw	2,001,709
DLmw	1,203,382

Using the data from Table 1, the I1, D1, and LL caches can be seen. Referencing [20], first three rows give us I cache, next three D cache reads and finally the last three tell the D cache writes. D1mr and D1mw give total D1 cache. LL cache is given by ILmr, DLmr and DLmw.

The Callgrind profiling tool is a call-graph generating cache profiler. Its output can be viewed using KCacheGrind tool. Output is shared in Table 2.

Table 2. Callgrind profiling

Incl.	Self	Called	Function
99.86	0.00	1	Main
99.86	0.00	1	Py_Main
98.85	0.00	1	PyRun_AnyFileExFlags
98.85	0.00	1	PyRun_SimpleFileExFlags
98.85	0.00	1	PyRun_FileExFlags
47.31	0.00	352	cv::parallel_for_(cv::Range const&,cv::ParallelLoopBody const&, double)
26.48	0.00	198	cv::MorphologyRunner::operator()(cv::Range const&) const
26.44	0.00	198	Cv::FilterEngine::apply
26.34	0.65	198	Cv::FilterEngine::proceed
20.13	0.00	66	Cv::medianBlur
19.57	0.00	22	Cv::Resize
19.47	17.47	22	Cv::ResizeArea_Invoker < unsigned char,float > ::operator()
17.89	0.00	132	Cv::erode
17.55	0.00	66	Cv::morphologyEx
13.67	0.00	23	Cv::VideoCapture::read
12.37	0.00	23	cvGrabFrame
12.37	0.65	23	CvCaptuereCam_V4L_CPP::grabFrame()
11.72	0.00	86	V4l2_iocti
11.71	0.00	23	V4lconvert_convert
10.63	10.63	29,832	Cv::MorphRowFilter
8.60	0.00	66	Cv::dilate

From Table 2, it can be concluded that OpenCV classes MorphologyRunner & FilterEngine get a significant amount of time being spent on them. Both are being called during morphological operations. The medianBlur class called to apply Median Filtering is also called many times. The percentage of time being spent on the Resize class is also large considering that the times it is being called is fewer.

4 Conclusions

The objective of recognizing traffic lights has been fulfilled with this method. There are hindrances which are faced like brightness of LED and other traffic signal like looking objects on road. To solve the problem of LED, shutter speed of camera can be changed. Another way to solve the problem of traffic light signal recognition could be use object detection algorithm based on [21] using Haar cascades. Distance estimation from the signal as performed in [22] can also be implemented.

Using deep learning algorithms like convolutional networks is also known to be beneficial. Convolutional neural networks are a type of feed-forward neural networks proposed by Alex Krizhevsky in [23]. These CNNs can be used to solve the problem of traffic light recognition as performed in [7, 24].

Work on this RC car can further be extended to make it more autonomous. Wireless modules can be installed making it properly mobile. Traffic signs recognition can be implemented using convolutional neural networks as discussed in [25]. To make the car turn by itself, a neural network can be trained to classify captured image road between going straight, curving left or curving right. This has been performed in [26] where a two-layered neural network has been trained for this purpose.

References

1. Jensen, M.B., Philipsen, M.P., Møgelmose, A., Moeslund, T.B., Trivedi, M.M.: Vision for looking at traffic lights: issues, survey, and perspectives. In: IEEE Transactions on Intelligent Transportation Systems (2015)
2. Williams, C.: Stop lights, sunsets, junctions are tough work for Google's robo-cars (2016). https://www.theregister.co.uk/2016/08/24/google_self_driving_car_problems/
3. Lee, U., et al.: EureCar turbo: a self-driving car that can handle adverse weather conditions. In: 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Daejeon, pp. 2301–2306 (2016)
4. Yu, C., Huang, C., Lang, Y.: Traffic light detection during day and night conditions by a camera. In: IEEE 10th International Conference on Signal Processing Proceedings, Beijing, pp. 821–824 (2010)
5. Campbell, J., Amor, H.B., Ang, M.H., Fainekos, G.: Traffic light status detection using movement patterns of vehicles. In: 2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC), Rio de Janeiro, pp. 283–288 (2016)
6. Shi, Z., Zou, Z., Zhang, C.: Real-time traffic light detection with adaptive background suppression filter. IEEE Trans. Intell. Transp. Syst. **17**(3), 690–700 (2016)

7. John, V., Yoneda, K., Liu, Z., Mita, S.: Saliency map generation by the convolutional neural network for real-time traffic light detection using template matching. IEEE Trans. Comput. Imaging **1**(3), 159–173 (2015)
8. Michael, M., Schlipsing, M.: Extending traffic light recognition: efficient classification of phase and pictogram. In: 2015 International Joint Conference on Neural Networks (IJCNN), Killarney, pp. 1–8 (2015)
9. Suzuki, S., Abe, K.: Topological structural analysis of digitized binary images by border following. Comput. Vision Graph. Image Proc. **30**(1), 32–46 (1985). http://dx.doi.org/10.1016/0734-189X(85)90016-7, ISSN 0734-189X
10. Fisher, R., Perkins, S., Walker, A., Wolfart, E.: Mathematical morphology. http://homepages.inf.ed.ac.uk/rbf/HIPR2/matmorph.htm
11. Najman, L., Talbot, H.: Mathematical morphology, 1st edn. ISTE, London (2010)
12. Fisher, R., Perkins, S., Walker, A., Wolfart, E.: Morphology. http://homepages.inf.ed.ac.uk/rbf/HIPR2/morops.htm
13. Tutorial: mathematical morphology. https://clouard.users.greyc.fr/Pantheon/experiments/morphology/index-en.html
14. Chanda, B., Majumder, D.: Digital image processing and analysis, 1st edn. Prentice-Hall of India, New Delhi (2007)
15. Fisher, R., Perkins, S.,Walker, A., Wolfart, E.: Median filter. http://homepages.inf.ed.ac.uk/rbf/HIPR2/median.htm
16. Huang, T., Yang, G., Tang, G.: A fast two-dimensional median filtering algorithm. IEEE Trans. Acoust. Speech Signal Process. **27**(1), 13–18 (1979)
17. Wang, R.: Median Filter. http://fourier.eng.hmc.edu/e161/lectures/smooth_sharpen/node2.html
18. Median Filtering. http://users.ecs.soton.ac.uk/msn/book/new_demo/median/
19. Remove noise using an averaging filter and a median filter. https://www.mathworks.com/examples/image/mw/images-ex74217292-remove-noise-using-an-averaging-filter-and-a-median-filter
20. Valgrind. http://valgrind.org/docs/manual/cg-manual.html
21. Viola, P., Jones, M.: Robust real-time face detection. In: Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001, pp. 747–747 (2001)
22. Diaz-Cabrera, M., Cerri, P., Sanchez-Medina, J.: Suspended traffic lights detection and distance estimation using color features. In: 2012 15th International IEEE Conference on Intelligent Transportation Systems, Anchorage, AK, pp. 1315–1320 (2012)
23. Krizhevsky, A., Sutskever, I., Hinton, G. E.: ImageNet classification with deep convolutional neural networks. In: Proceedings of NIPS, pp. 1097–1105 (2012)
24. Philipsen, M.P., Jensen, M.P., Møgelmose, A., Moeslund, T.B., Trivedi., M.M.: Learning based traffic light detection: evaluation on challenging dataset. In: 18th IEEE Intelligent Transportation Systems Conference (2015)
25. Li, Y., Møgelmose, A., Trivedi, M.M.: Pushing the "Speed Limit": high-accuracy US traffic sign recognition with convolutional neural networks. IEEE Trans. Intell. Veh. **1**(2), 167–176 (2016)
26. Wang, Z.: Self driving RC car. https://zhengludwig.wordpress.com/projects/self-driving-rc-car/
27. Online diagram software to draw flowcharts, UML & more. Creately. http://www.creately.com

RGB-Depth Image Based Human Detection Using Viola-Jones and Chan-Vese Active Contour Segmentation

Puji Lestari[1,2](✉) and Hans-Peter Schade[1]

[1] Institut Für Medientechnic, Technische Universität Ilmenau,
Ilmenau, Germany
{puji.lestari,schade}@tu-ilmenau.de
[2] Research Center for Informatics, Indonesian Institute of Sciences,
Jakarta, Indonesia

Abstract. Human detection refers to the process of detecting human region from an image or from video frames. Most of the recent advanced human detection systems use the segmentation scheme by incorporating the depth information of the scene. In such systems, the scene gets captured by a RGB-D camera and the candidate area is segmented by setting an appropriate depth threshold for the captured depth images. In practice, depth data obtained from this depth analysis having critical problems, such as optical noise, absence of depth information for certain regions like hair area, and unmatched boundaries. The proposed approach mainly focus on restoring the actual edge information and hair area of the subject in the pre-segmented image by applying Viola Jones Algorithm for face area detection and Chan-Vese active contour detection for restoring hair and edge areas of the image over the detected face area. This final segmentation mask is used for segmenting the accurate human region from the original image with hair area and with boundaries similar to the ground truth. Experimental results prove the improvement in the visual quality of the segmented human area.

Keywords: Kinect · Depth analysis · Active contour detection

1 Introduction

Understanding human motion and activity tracking from a video is important in numerous applications, especially in video surveillance systems [1–3]. The main stage for understanding human motion from a video is human detection and segmentation [4]. However, detection and tracking of human from images or videos is a challenging problem facing by researchers due to variations in pose, clothing, lighting conditions and vivid backgrounds. There has been ample research [5, 6] in the recent years in human detection and various methods based on computer vision have been proposed.

Basic conventional human detection approaches is usually done in images taken by cameras which gives RGB color images [7]. These methods adopt the detection process by analyzing various features of color and shape. Such systems use features based on gradients, such as histograms of oriented gradients (HOG) [8], or extract interest points

S.M. Thampi et al. (eds.), *Advances in Signal Processing and Intelligent Recognition Systems*, Advances in Intelligent Systems and Computing 678,
DOI 10.1007/978-3-319-67934-1_25

in the image, such as scale-invariant feature transform (SIFT) [9]. Such methods follow a similar computer vision approach as that of human vision system. Even though lots of reports [10, 11] showed that these methods can provide sufficient human detection results, RGB image based methods faces difficulties in perceiving the extract shapes of the human subjects with articulated poses or when the background is cluttered. Background subtraction techniques [12] usually focus on such color based features along with motion detection analysis which helps to identify foreground moving human subjects from video frames. This approach operate directly on the image or video frames to classify moving objects as human or non-human using shape, color, motion features or combinations of these.

There has been research using range image [13] for object recognition or modeling in the past few decades. Range images have many advantages over normal 2D Color or Gray intensity images. Range images are robust to the change in color and illumination. Also, range images can give simple representations of 3D information. Recently, many human detection methods using depth information taken by the RGB-D camera has been proposed. RGB-D camera can capture both color and depth information from an image, where the depth represents the distance between objects and the camera. These cameras have been used and cheap and easy to use versions of these sensors are available like Microsoft Kinect [14, 15], Intel Realsense [16] and Asus Xtion [17]. Depth information is a crucial cue for detecting objects in cases where the objects may not have consistent color and texture. Hence, human detection using RGB-D images has many potential applications in future techs including smart car, Smart visual surveillance, human-computer interaction and robotic navigation.

Many human detection models [15–18] are proposed recently based on depth and RGB color information. Their main disadvantages include optical noise, absence of depth information for certain regions like hair area, unmatched contour detection etc. In this work, we attempt to address the problem of human detection from video frames or from still images without any loss in hair area and to reduce distortion over the edge areas. Microsoft Kinect camera is used image capturing to get the depth and color information. This sensor (camera) can take depth information and color information simultaneously using the same optical axis, thereby achieving 3D detection by a single camera. The proposed system is performed over regions obtained from a pre-segmentation of the depth image. The head portion get first tracked by using Viola-Johnson Face detection algorithm [19, 20] and then Chan-Vese active contour segmentation [21] method is applied to restore the lost hair portion.

The rest of the paper is organized as follows. Section 2 contains a brief survey of recent related works in human detection which uses depth analysis. Section 3 contains a detailed description of proposed work methodology. Discussion about the experimental results is plotted in Sect. 4. Finally, the concluding remarks are drawn in Sect. 5.

2 Literature Survey

Human detection from depth images is gaining substantial attention since depth information facilitates object extraction and from a relative distant background. In most of the previous works, dealing with human detection using depth analysis, the primary objective is to detect and track moving or stationary humans. In some of the methods developed to detect human candidates from individual depth frames, it required the ground or the ceiling is visible. In the algorithm proposed by Bagautdinov et al. [22] searched for the presence of humans on a ground surface using Bayesian inference. In this paper, a method is proposed to track moving especially walking people from RGB-Depth images. The ground plane is detected first and removed from the point cloud data, and the remaining points are clustered to determine the human region.

In another article, Xia et al. [23] developed a human detection approach which aims at first detecting head areas from the edge map of depth images. They use Chamfer distance to a template head contour to find subject's head region, then verify using a 3D head sphere model. To decrease the high false positive rate, association between successive video frames is also used. Choi et al. [24] employed a graph-based segmentation algorithm followed by a region merging operation to determine candidate human factor. They use linear SVM to classify the human area by computing Histogram of Depth (HOD) descriptors. W. Choi et al. [25] developed a human tracking system that applies various cues and detectors to video data. Depth information is integrated in this system whenever it is available. Their depth-based shape detector employs a binary head-and-shoulder template to calculate the likelihood of human presence in target locations.

In the paper [26], human detection algorithm is composed of three main steps: A pre-segmentation stage of the depth scene using K-means clustering [27], and merging of adjacent planar regions. The second step involves the extraction of omega like curves from top portions of boundaries of the segmented area, and matching them with template head-shoulder curves. Finally, the candidate head-shoulder regions are inspected to verify whether they satisfy two geometrical constraints attributed to valid head and shoulder regions. Another similar methodology is proposed in the article which detects moving objects in depth image sequences using background images and motion depth (MD) analysis [28]. The background image represents the camera view with no moving objects and the MDs are the depth values corresponding to moving objects. Foreground regions are then isolated and detected by background subtraction. Experimental results show the proposed method robustly detects moving objects, even if the moving region exists in close proximity to the background region. Therefore, show better performance in detecting moving objects including human subjects in depth image sequences.

The most common issue found in almost all above said methods is that, the segmentation by depth analysis alone is not sufficient to extract the complete details of the human area. However, the depth camera does not capture shiny and dark surfaces well, such as black hair since the reflected lights from the dark and shiny surfaces are weak and scattered, the depth camera cannot detect the reflected infrared light. For example depth analysis usually fails to segment hair area from the face region since it is usually seen in black or dark color. This issue doesn't bring any problem if the objective is just detection and tracking. But if the intention of the analysis is to extract human region for applications like Animation or Graphics purpose, the boundary details and the hair portion detection will be necessary.

A similar issue is pointed in the paper [29], where the authors propose a new method of generating a dynamic 3D human actor using a Time of Flight (TOF) depth camera.

The solutions suggested here could minimize the problems inherent in the depth camera system. They detect the lost hair region using a computer vision based face detection technique and recover its depth information using a multi-seed region growing algorithm considering the depths of the face region. Then they match the boundaries between the color and depth images using a graph cut-based matting algorithm. Another attempt of human recognition is explained in the paper [30] where proposing a method for detecting humans by Relational Depth Similarity Features based on depth data from a TOF camera. The used features are derived from a similarity of depth-histograms which represent the relationship between two local regions. During the detection process, a raster scanning in a 3D space is used which makes the detection is a faster way. A considerable increase in speed is achieved.

3 Proposed Method

The Proposed system aims at developing an efficient post processing on pre segmented regions of people after depth analysis. The system aims at restoring the hair portion that get lost in the depth based segmentation process. The subsequent steps of the proposed algorithm are shown in Fig. 1.

3.1 Pre-segmentation of Human Region Using Depth Analysis

The main objective of this work is to enhance a vague pre-segmented image obtained after processing the RGB-Depth image using color and depth analysis. These images may

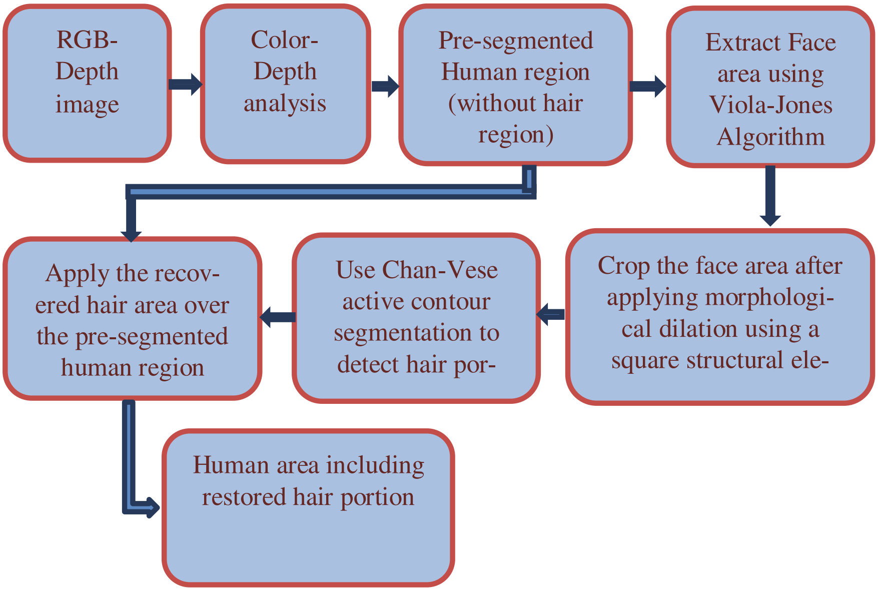

RGB-Depth image
Color-Depth analysis
Pre-segmented Human region (without hair region)
Extract Face area using Viola-Jones Algorithm
Crop the face area after applying morphological dilation using a square structural ele-
Use Chan-Vese active contour segmentation to detect hair por-
Apply the recovered hair area over the pre-segmented human region
Human area including restored hair portion

Fig. 1. Block diagram of the proposed method.

be captured by camera like Microsoft Kinect [14]. Since we are focusing the enhancement over the initial segmented image by restoring the lost hair region [29], the following techniques are being used and explained below. A sample of the pre-segmented image result after the color- depth feature based segmentation process is shown in Fig. 2.

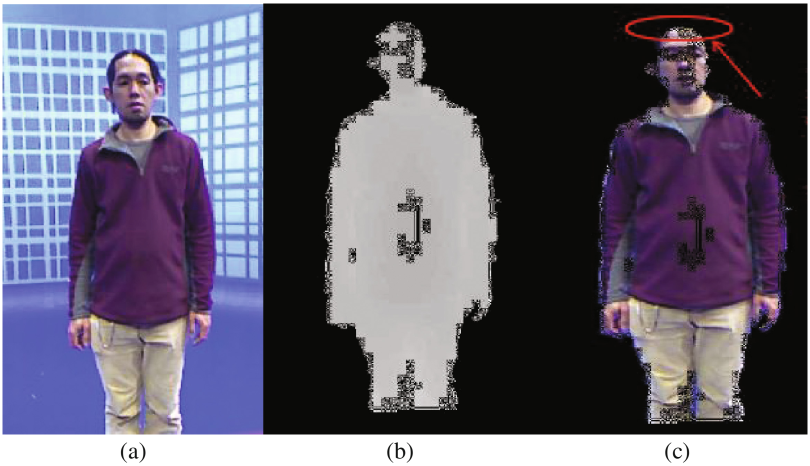


Fig. 2. (a) RGB image (b) depth image & (c) segmented human area using depth analysis

3.2 Face Area Detection Using Viola-Jones Algorithm

The objective of the proposed system is restoring the hair area from the course segmented image. The system adopts a way to detect the face region first in order to do the same. The main issue to be solved here is the effective detection of face region from the image. During the last decade a number of face detection algorithms have been developed. Among these Viola-Jones algorithm is used here while considering its performance over various advantages [31, 32].

The basic idea of the Viola-Jones algorithm [20] is to scan a sub-window capable of detecting faces across a given input image. The standard image processing approach would be to rescale the input image to different sizes and then run the fixed size detector through these images. But Viola-Jones has devised a scale-invariant detector that needs the same number of calculations irrespective of the size of the image. This detector is constructed using a so-called integral image and some simple rectangular features reminiscent of Haar-wavelets. The detected portion of face using Viola-Jones algorithm is represented in Fig. 3.

Morphological dilation [33] with a square structural element is applied to the face mask area in order to cover the entire portion around the face in order to include the hair region. The further active contour based segmentation [21] is applied on this cropped face area.

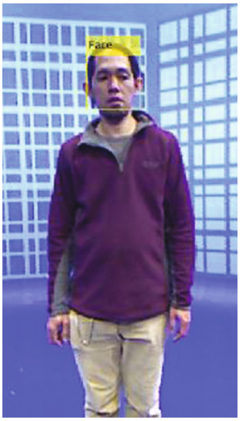


Fig. 3. Face detection using viola jones

3.3 Chan-Vese Active Contour Segmentation

Here we use Chan-Vese active contour segmentation method to segment the hair region from the cropped face area. The method optimally fits a two-phase piecewise constant model on the cropped face image. The segmentation contour is represented indirectly using a level set function [34], which allows the segmentation more easily than explicit snake methods [35]. Usually, the hair region and face skin areas are statistically different and homogeneous; this algorithm gives much better results. The approach based on the minimization of an energy based-segmentation. The important stages used in this step are explained here.

Let the pixel value at any given position (x, y), in the cropped face image is denoted by $u_0(x, y)$, the average value inside C_1 is denoted by c_1, and average value outside the region C_2 is denoted by c_2. In Chan-Vese Segmentation algorithm, the following fitting term is considered

$$F_1(C) + F_2(C) = \int\limits_{C_1} |u_0(x, y) - c_1|^2 dxdy + \int\limits_{C_2} |u_0(x, y) - c_2|^2 dxdy \qquad (1)$$

From the Eq. (1), the evolved curve is at the boundary of an object when the sum of $F_1(C)$ and $F_2(C)$ is approximately zero. The first term of the summation indicates the criteria that, when the curve lies inside the object, its value will be closer to zero and its value will be greater than zero when the curve is outside the object area. Similarly, the value of the second term will be approximately zero when the curve is outside the object and greater than zero when it lies inside the object. Therefore at the point while getting the summation of both of the terms as zero, the curve might be around the edge which separates the foreground and background.

By including two additional terms; the curve length and area inside the curve, minimization of the Eq. (1) is improved. In addition, the parameters μ, v, λ1, and λ2 are added to the equation to give proper weight to the contribution of each term to the final energy function equation of the reference method. Now, the energy function is rewritten as:

$$\begin{aligned} F_1(c_1, c_2, \varnothing) =& \mu.Length(C) + \vartheta.Area(C_1) + \lambda_1 \int\limits_{C_1} |u_0(x,y) - c_1|^2 dxdy \\ & + \lambda_2 \int\limits_{C_2} |u_0(x,y) - c_2|^2 dxdy \end{aligned} \tag{2}$$

A level set formulation is then applied for the minimization of energy function, where C in Ω is represented by the zero level set of a Lipchitz function φ: $\Omega \rightarrow$ R, such that

$$\begin{aligned} C &= \partial\omega = \{(x,y) \in \Omega : \varnothing(x,y) = 0\}, \\ C_1 &= \omega = \{(x,y) \in \Omega : \varnothing(x,y) > 0\}, \\ C_2 &= \omega/\bar{w} = \{(x,y) \in \Omega : \emptyset(x,y) < 0\}, \end{aligned} \tag{3}$$

Thus the final version of the energy function is expressed as:

$$\begin{aligned} F_1(c_1, c_2, \varnothing) =& \mu \int\limits_{\Omega} \delta(\varnothing(x,y)) |\nabla \varnothing(x,y)| dxdy \\ & + \vartheta \int\limits_{\Omega} H(\varnothing(x,y)) dxdy \\ & + \lambda_1 \int\limits_{\Omega} |u_0(x,y) - c_1|^2 H(\varnothing(x,y)) dxdy \\ & + \lambda_2 \int\limits_{\Omega} |u_0(x,y) - c_2|^2 H(\varnothing(x,y)) dxdy \end{aligned} \tag{4}$$

Here we chose $\lambda_1 = \lambda_2 = 1$, $\vartheta = 0$ and $\mu = .2$ for classifying the hair region from the face skin area. The segmented foreground area (hair region) is then merged with the

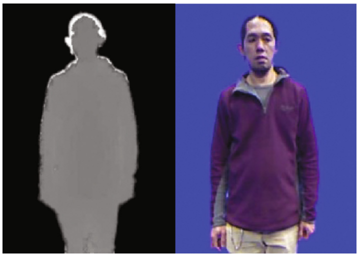


Fig. 4. (a) Final mask for human detection, (b) detected human area

earlier pre-segmented mask and the resultant mask image is shown in Fig. 4(a). The white area in the mask represents the detected hair region using Chan-Vese active contour segmentation. The enhanced mask is used to segment the human region from the image/video frame and the segmented human area is shown in Fig. 4(b).

4 Experimental Results

Here for analyzing and evaluating the performance of Hair region recovery and corresponding enhancement in segmentation mask, we used custom datasets of RGB-Depth videos taken by Microsoft Kinect camera under constrained environments. The color (RGB) and depth videos are taken simultaneously by using the Kinect sensor and saved separately. The videos were recorded at 480 x 640 resolutions and with a frame rate of 5 fps to reduce the unnecessary computational complexity. A sample frame and the various results obtained are shown below.

A Precision, Recall and accuracy based analysis is used to compare the performance of the proposed system with similar existing techniques. These performance parameters were defined as in Eq. (5).

$$\begin{aligned} Precision &= \frac{TP}{TP+FP} \\ Recall &= \frac{TP}{TP+FN} \\ Overall\,Accuracy &= \frac{TP+TN}{TP+FP+TN+FN} \end{aligned} \tag{5}$$

Where TP is the True Positive, TN is True Negative, FP is the False Positive and FN is the False Negative pixels segmented during the segmentation process. Table 1 gives the performance analysis of the proposed method with other methods.

From the results, it is obvious that the proposed enhancement in segmentation mask helps to recover the human area more effectively. The computation time required is slightly higher than a compared scheme but the segmentation quality is far better than other approaches. From the sample output images as shown in Fig. 5, the boundary

Table 1. Performance analysis

Method	Precision (%)	Recall (%)	Overall accuracy (%)	Average computaion time (in s)
Proposed scheme	99.2	93.6	95.3	5.57
Xia et al. [23]	93.5	82.4	88.4	9.5
Choi et al. [24]	92.5	73.7	82.3	6.23
Ikemura et al. [30]	90.4	62.9	78.5	4.91

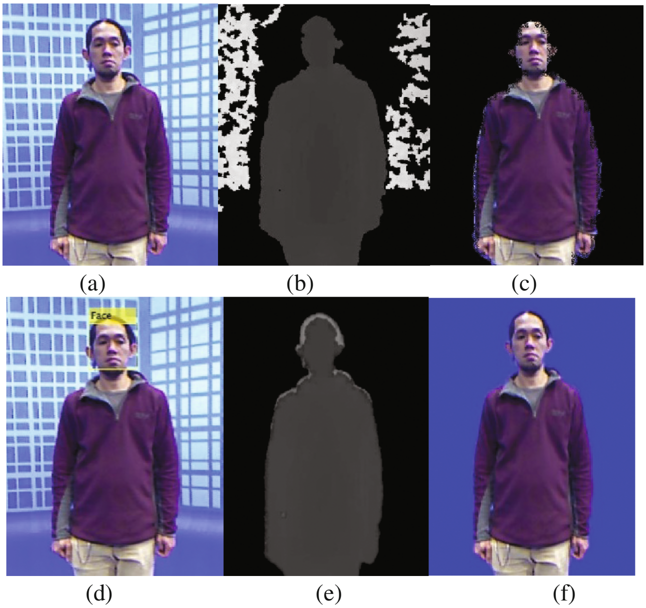


Fig. 5. (a) RGB frame,(b) depth Image, (c) pre-segmented human area, (d) detected face region using Vialo-Jones algorithm, (e) enhanced segmentation mask (white region shows the recovered hair area using proposed method), (f) Segmented human region using the enhanced mask

regions become smoother and more effective while using the proposed method. The hair area recovery helps to enhance both the statistical and aesthetic quality of the segmented results.

5 Conclusion and Future Work

The proposed method is suitable for the accurate and effective recovery of lost hair region from pre-segmented raw depth images. Segmentation masks for extracting human region from scene by utilizing Chan-vese active contour segmentation. The approach utilizes an unsupervised segmentation process and is devoid of training time. From the experimental results, it can be seen that the enhancement of the segmentation mask is considerably good and the boundaries are much clear and accurate, while compared to the pre-segmented image. In future research, the proposed algorithm may be modified to apply the enhancement over unconstrained video samples and reliable human detection and tracking based applications can be developed.

References

1. Dhamsania, C.J, Ratanpara, T.V.: A survey on human action recognition from videos. In: 2016Online International Conference on Green Engineering and Technologies (IC-GET), Coimbatore, pp. 1–5 (2016). doi:10.1109/GET.2016.7916717
2. Chin-Teng, L., Linda, S., Yu-Wen, S., Tzu-Kuei, S.: A conditional entropy-based independent component analysis for applications in human detection and tracking. EURASIP J. Adv. Sig. Proc., 468329 (2010)
3. Jardim, D., Nunes, L., Dias, M.: Human activity recognition from automatically labeled data in RGB-D videos. In: 2016 8th Computer Science and Electronic Engineering (CEEC), Colchester, pp. 89–94. (2016). doi:10.1109/CEEC.2016.7835894
4. Zaihidee, E.M., Ghazali., A.A, Almisreb, K.H.: Comparison of human segmentation using thermal and color image in outdoor environment. In: 2015 IEEE Conference on Systems, Process and Control (ICSPC), Bandar Sunway, pp. 152–156 (2015). doi:10.1109/SPC.2015.7473576
5. Bilinski, P., Bremond, F.: Human violence recognition and detection in surveillance videos. In:2016 13th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), Colorado Springs, CO, pp. 30–36. (2016) doi:10.1109/AVSS.2016.7738019
6. Xinjian, Z., Liqing, Z.: Real Time Crowd Counting with Human Detection and Human Tracking. In: International Conference on Neural Information Processing, ICONIP: Neural Information Processing, pp. 1–8 (2014)
7. Tejero-de-Pablos, A., Nakashima, Y., Sato, T., Yokoya, N.: Human action recognition-based video summarization for RGB-D personal sports video. In: 2016 IEEE International Conference on Multimedia and Expo (ICME), Seattle, WA, pp. 1–6 (2016). doi:10.1109/ICME.2016.7552938
8. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2005), San Diego, CA, USA, vol. 1, pp. 886–893 (2005)
9. David, G.L.: Distinctive Image Features from Scale-Invariant Keypoints. Computer Science Department University of British Columbia Vancouver, B.C., Canada
10. Singh, S., Gupta, S.C.: Human object detection by HoG, HoB, HoC and BO features. In: 2016 Fourth International Conference on Parallel, Distributed and Grid Computing (PDGC), Waknaghat, pp. 742–746 (2016). doi:10.1109/PDGC.2016.7913220
11. Khawlah, H.A., Tianjiang W.: Recognition of human action and identification based on SIFT and Watermark. In: International Conference on Intelligent Computing ICIC: Intelligent Computing Methodologies, pp. 298–309 (2014)
12. Thakore, D.G., Trivedi, A.I., : Prominent boundaries and foreground detection based technique for human face extraction from color images containing complex background. In: 2011 Third National Conference on Computer Vision, Pattern Recognition, Image Processing and Graphics, Hubli, Karnataka, pp. 15–20 (2011). doi:10.1109/NCVPRIPG.2011.11
13. Zhang, B., Liu, Q., Ikenaga, T.: Ghost-free high dynamic range imaging via moving objects detection and extension. In: 2015 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA), Hong Kong, pp. 459–462 (2015). doi:10.1109/APSIPA.2015.7415313
14. Procházka, A., Martin, S., Oldˇrich, V., Martin, V.: Microsoft kinect visual and depth sensors for breathing and heart rate analysis. Sensors **16**, 996 (2016). doi:10.3390/s16070996

15. Stone, E.E., Skubic, M.: Fall detection in homes of older adults using the Microsoft Kinect. IEEE J. Biomed. Health Inform. **19**(1), 290–301 (2015). doi:10.1109/JBHI.2014.2312180
16. Anggraini, N., Rozy, N.F., Lazuardy, R.A.: Facial recognition system for fatigue detection using Intel Realsense Technology. In: 2016 6th International Conference on Information and Communication Technology for The Muslim World (ICT4 M), Jakarta, pp. 248–253 (2016). doi:10.1109/ICT4M.2016.058
17. Walas, K., Nowicki, M., Ferstl, D., Skrzypczyński, P.: Depth data fusion for simultaneous localization and mapping — RGB-DD SLAM. In: 2016 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI), Baden-Baden, pp. 9–14 (2016). doi:10.1109/MFI.2016.7849459
18. Cao, Y., Shen, C., Shen, H.T.: Exploiting depth from single monocular images for object detection and semantic segmentation. IEEE Trans. Image Process. **26**(2), 836–846 (2017). doi:10.1109/TIP.2016.2621673
19. Viola, P., Jones, M.: Robust real-time face detection. In: Eighth IEEE International Conference on Computer Vision. ICCV 2001, p. 747 (2001). doi:10.1109/ICCV.2001.937709
20. P, Viola., M, Jones.: Rapid object detection using a boosted cascade of simple features. In: Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2001, vol. 1, pp. I-511–I-518 (2001)
21. Chan, T.F., Vese, L.A.: Active contours without edges. IEEE Trans. Image Process. **10**(2), 266–277 (2001). doi:10.1109/83.902291
22. Bagautdinov, T.M., Fleuret, F., Fua, P.: Probability occupancy maps for occluded depth images. In: IEEE Conference on Computer Vision and Pattern Recognition. CVPR 2015, Boston, MA, USA, 7–12 June 2015, pp. 2829–2837 (2015)
23. Lu, X., Chia-Chih, C., Aggarwal, J.K.: Human detection using depth information by Kinect. In: CVPR 2011 WORKSHOPS, Colorado Springs, CO, pp. 15–22 (2011). doi:10.1109/CVPRW.2011.5981811
24. Benjamin, C., Cetin, M., Joydeep, B., Manuela, V.: Fast human detection for indoor mobile robots using depth images. In: 2013 IEEE Internationsal Conference on Robotics and Automation (ICRA) Karlsruhe, Germany, 6–10 May 2013
25. Choi, W., Pantofaru, C., Savarese, S.: A general framework for tracking multiple people from a moving camera. IEEE Trans. Pattern Anal. Mach. Intell. **35**(7), 1577–1591 (2013)
26. Can, G.N., Dutagaci, H.: Human detection from still depth images. In: Society for Imaging Science and Technology. Image Processing, Measurement (3DIPM), and Applications, pp. 3DIPM-046.1–3DIPM-045.7 (2016)
27. Katkar, J., Baraskar, T., Mankar, V.R.: A novel approach for medical image segmentation using PCA and K-means clustering. In:2015 International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT), Davangere, pp. 430–435 (2015). doi:10.1109/ICATCCT.2015.7456922
28. Walia, G.S., Kapoor, R., Singh, S.: Depth and scale modeling of object for 3D motion analysis in video. In: 2013 IEEE Second International Conference on Image Information Processing (ICIIP-2013), Shimla, pp. 90–95 (2013). doi:10.1109/ICIIP.2013.6707562
29. Cho, J.H., Kim, S.Y., Ho, Y.S., Lee, K.H.: Dynamic 3D human actor generation method using a time-of-flight depth camera. IEEE Trans. Consum. Electron. **54**(4), 1514–1521 (2008). doi:10.1109/TCE.2008.4711195
30. Ikemura, S., Fujiyoshi, H.: Real-Time human detection using relational depth similarity features. In: ACCV 2010. LNCS, vol 6495, pp. 25–38 (2011)
31. Egorov, A.D., Shtanko, A.N., Minin, P.E.: Selection of Viola-Jones algorithm parameters for specific conditions. Bull. Lebedev Phys. Inst. **42**, 244 (2015). doi:10.3103/S1068335615080060

32. Bruce, B.R., Aitken, J.M., Petke, J.: Deep parameter optimisation for face detection using the Viola-Jones algorithm in OpenCV. In: Sarro, F., Deb, K. (eds.) Search Based Software Engineering, SSBSE 2016. LNCS, vol 9962. Springer, Cham (2016)
33. Gonzalez, C., Rafael, E., Woods, R.: Digital image processing, Prentice Hall India, 2002
34. Malladi, R., Sethian, J.A., Vemuri, B.C.: A topology independent shape modeling scheme. In: Proceeding of SPIE Conference Geometric Methods Computer Vision II, San Diego, CA, vol. 2031, pp. 246–258 (1993)
35. Kass, M., Witkin, A., Terzopoulos, D.: Snakes: active contour models. Int. J. Comput. Vis. **1**, 321–331 (1988)

Choice of the Scheduling Technique Taking into Account the Subcontracting Optimization

Konstantin Aksyonov, Anna Antonova(✉), and Natalia Goncharova

Ural Federal University, Mira 19, Ekaterinburg, Russia
Wiper99@mail.ru, antonovaannas@gmail.com,
n.v.goncharova@urfu.ru

Abstract. This paper analyzes a number of scheduling technique from classical network planning techniques to hybrid techniques based on the integration of knowledge-based simulation, evolutionary modeling, and heuristic search. The choice of the scheduling technique is based on the following criteria. First, solving the problems of subcontracting optimization, works rescheduling, analysis of alternative plans, renewable and non-renewable resources consideration. Secondly, application of the knowledge-based modeling and heuristic methods in the process of these problems solving. Taking into account the above criteria, a hybrid technique of multiagent genetic optimization has been chosen as the scheduling technique with subcontracting optimization.

1 Introduction

The subcontracting scheduling problem is part of the scheduling problem and is associated with the analysis of the bottlenecks of own resources distribution and optimisation of the volume and cost of subcontracting resources attracted to eliminate bottlenecks. The subcontracting scheduling problem is relevant especially for enterprises, where number of employees is small. The instability of the economic situation requires management to solutions aimed, on the one hand, to save money, but on the other hand, to maintain the company's competitiveness in the market. One of the obvious solutions is reducing the number of own employees and attraction subcontractor in the case of staff shortages. Subcontracting scheduling is aimed at the selection of a work schedule that would ensure maximum loading of own resources and execution of all projects on time due to additional resources. The objective function of subcontracting scheduling is to maximize the load of own resources and minimize the cost of subcontracting resources. Constraints of the problem are time constraints of the early and late works start. Let us consider a number of methods [2, 4–8] for solving the problem of work planning and carry out comparative analysis in relation to the problem of subcontracting scheduling.

S.M. Thampi et al. (eds.), *Advances in Signal Processing and Intelligent Recognition Systems*, Advances in Intelligent Systems and Computing 678,
DOI 10.1007/978-3-319-67934-1_26

2 Network Planning Techniques

Network planning techniques are based on the idea of presenting project work as a network. In this network, the aimed arcs are associated with the works, the nodes - with the events of the work's start and finish. A network schedule construction is carried out according to the rules and necessarily reflected the relationship between the preceding and following works because the performance of the network planning techniques is based on the analysis of these relationships.

Let us consider the following network planning techniques: critical path method (CPM), program evaluation and review technique (PERT), and graphical evaluation and review technique (GERT).

2.1 CPM Method

The CPM method is intended to assess the standby time of works performance in the case of deterministic works durations. The found standby time of works performance is used to balance resource, which is carried out through a variety of heuristic algorithms to establish the priority of works [5]. After resource balancing is finished, the only way to reduce the critical path is to attract subcontracting resources. To assess the feasibility of attracting subcontracting resources, a value of the average cost of reducing the duration of project per unit time is calculated for each work. Then the most profitable work from the viewpoint of acceleration of project implementation is selected to attract subcontracting resources.

2.2 PERT and GERT Methods

The PERT method is intended to assess the timing of the project completion taking into account the assignment of the works duration by using β-distribution [5]. The GERT method is a development of the PERT method. The GERT method is intended to the analysis of stochastic network graph [6]. Each arc of the stochastic network (i.e. work) is characterized by the duration and the probability of realization in the project.

The network implementation is a network section, in which some arcs are stored (realized), while others arcs are discarded. Each node of the stochastic network is identified with two events: event of the work end (input event) and event of the work start (output event).

Three logical operations are defined in the GERT language for describing the input event: XOR (among all arcs, included in the node, only one arc can be realized), OR (any of arcs, included in the node, can be realized), AND (all arcs, included in the node, are to be implemented). Two types of the output are defined in the GERT language for describing the output event: deterministic output (all arcs, originating from the node, are implemented) and probabilistic output (only one arc of all the arcs, originating from the node, is implemented).

3 Scheduling Method on the Basis of the Needs-and-Means Network

Scheduling method on the basis of the needs-and-means network is proposed by Skobelev P.O. and is intended to schedule by constructing a multiagent system of the operative resource allocation in real time [8]. The method takes into account the possibility of adjusting the composition and characteristics of the planned projects, works and resources. The needs-and-means network (NM-network) is a multiagent system where each agent is characterized by needs (N) and means (M). The NM-network agent negotiates for meet their own needs with the help of other agent's means. In the method, the NM-network is used to represent the set of orders, projects, works and resources of the enterprise. Each item listed in the NM-network is associated with an agent, e.g., resource #1 agent, work #1 agent, project #1 agent, etc. During the agents negotiations, the distribution of resources by works is carried; fragment of the negotiations protocol at the level of individual agents is shown in Fig. 1.

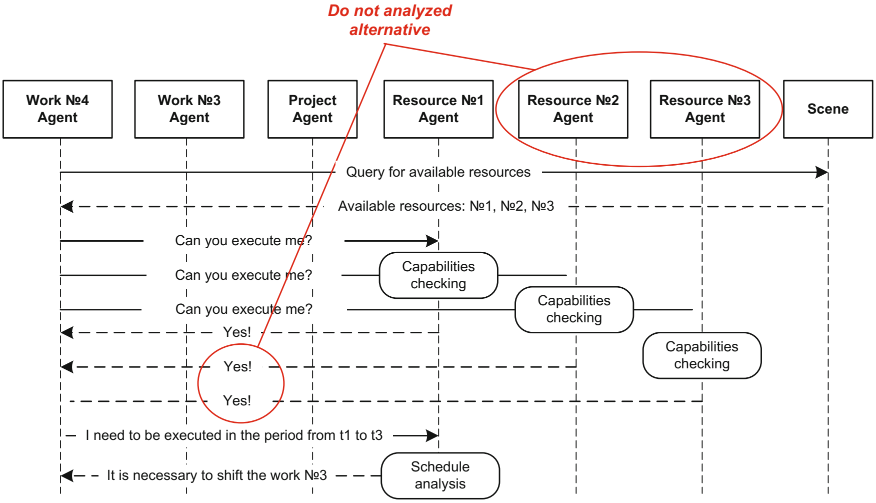


Fig. 1. A fragment of the negotiations protocol of the scheduling method on the basis of the NM-network

As can be seen from the figure, the work #4 agent chooses the resource#1 agent from the three available resource agents (the first in the list), so the other alternative resources remain unanalyzed.

Scheduling method on the basis of the NM-network includes an initial phase of the conflict-free scheduling and proactive phase of the rescheduling [7, 8].

The phase of the the conflict-free scheduling involves the resources allocation via successive selection by agents-work the necessary resources and generation the time interval (slot) of the resource employment, that is recorded in the planned slots of the agent-resource. The phase of the proactive scheduling requires resolution for each agent-resource the conflicts of the planned slots designated by agents-work. The conflicts resolution is performed by recursive negotiations according to the described agent's interaction protocol in the NM-network.

In the case of the new work or resource emergence, the system creates a new agent-work or agent-resource that is included in the negotiations, expressing its needs and means. Thus, the flexibility of the scheduling method is achieved in relation to changing environmental conditions.

4 Scheduling Method on the Basis of the Simulation and Genetic Algorithm Integration

Scheduling method on the basis of the simulation and genetic algorithms integration is proposed by Kureichik V.V. Fig. 2 shows a diagram of the Kureichik V.V. method.

The Kureichik V.V. method is intended to modeling discrete processes that occur in the organizational and technical systems, and optimisation of control process parameters by means of genetic algorithms (GA). Genetic algorithms are widely known as algorithms for solving the complex systems management problem in a short time [4].

According to the Kureichik V.V. method, the scheduling problem is solved by the GA, and the simulation model is used to calculate the multiobjective function of the suitability of individuals of the next generation. Expert system is used to analyze and correct parameters of the GA (the probability of genetic operators).

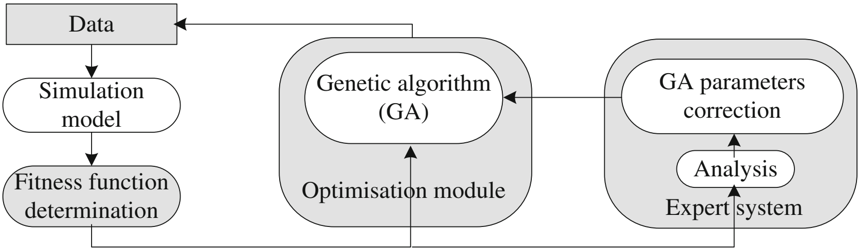


Fig. 2. A diagram of the Kureichik V.V. method

The proposed method has been implemented in the software RDO-studio [4] and with his help the scheduling problem of the shop works has been solved. The purpose of optimisation was the selection of the optimal values of the controlled variables of the simulation model, which were the priorities of executive works. Fitness function was the function of penalties for failures orders.

5 Method of Multiagent Genetic Optimisation

The method of multiagent genetic optimisation (MGO-method) has been developed by the paper authors [2]. This method is intended to subcontracting scheduling subject to possible portfolio dimension change. At the heart of the MGO-method is a Kureichik V.V. method modified using the following methods and algorithms: (1) simulation for description of the model of distribution of internal and external resources on works; (2) multiagent knowledge-based modeling (with production rules as knowledge representation model) for decision-makers behaviour description; (3) numerical methods of uncertainty removing with regard to the account the probability of occurrence of additional projects; (4) simulated annealing algorithm and search of novelty algorithm in order to modification of the genetic algorithm (GA) to improve the quality of solutions found by the algorithm. The MGO-method is implemented in the optimisation module of the metallurgical enterprise information system (MEIS). The MEIS is a web-oriented knowledge-based decision support system for tracking, monitoring, modeling, and analysis processes of the steel products manufacturing [3].

During solving the subcontracting problem under uncertainty, the following steps are carried out: (1) multiagent system model development; (2) forming a set of environment states; (3) forming an initial population of the modified GA; (4) for each chromosome of the population: (4.1) launch the simulation model with the adjustable parameters that are equal to phenotype of the chromosome, (4.2) chromosome's fitness function evaluation by output model parameter value; (5) forming next population by genetic operators applying; (6) repeat steps 4 and 5 until GA stopping criterion true; (7) uncertainty removing by replacement the objective functions vector at scalar quantity; (8) better decision choosing. The criterion for GA stopping is achieving a specified number of chromosomes populations.

6 Comparative Analysis of the Subcontracting Scheduling Techniques

Let us consider the following subcontracting scheduling techniques: network planning techniques (NPT), scheduling method on the basis of the needs-and-means network (NMN), scheduling method on the basis of the simulation and genetic algorithms integration (SGA), and MGO-method (MGO). Results of these techniques comparative analysis are presented in Table 1.

As follows from the table, all of the methods, except the MGO-method, do not have the full functionality of the subcontracting scheduling.

1. The bulkiness and poor readability of the GERT stochastic network diagrams. During conversion any stochastic network to this type, the number of arcs and nodes increases, which complicates the perception of the network.
2. Absence of the subcontracting optimisation techniques. The CPM method allows assigning third-party resources to perform activities of the critical path.
3. Lack of means of formalizing the decision-making scenarios in the allocation of resources on the works.

Table 1. Comparative analysis of the subcontracting scheduling techniques

Criterion	NPT	NMN	SGA	MGO
Problems				
Scheduling	•	•	•	•
Renewable resources consideration	•	•	•	•
Non-renewable resources consideration	○	○	○	•
Subcontracting consideration	•	•	○	•
Subcontracting optimisation	○	○	○	•
Rescheduling	○	•	○	•
Analysis of alternative plans	○	○	•	•
Methods for solving the scheduling problem				
Simulation	○	○	•	•
Multagent modeling	○	•	○	•
Evolutionary modeling	○	○	•	•
Knowledge-based modeling	○	•	•	•
Heuristic algorithms	•	○	•	•

NMN-method uses the negotiations of intelligent agents via knowledge-based modeling, but these agents do not support the analysis of alternative solutions, cutting off the "extra" alternatives in the course of negotiations.

Analysis of the SGA-method reveals the following disadvantages:

1. Method focuses on a wide class of the organizational and technical systems management tasks, which leads to the need to develop custom ontology of the subcontracting scheduling problem and develop of its own genetic algorithm.
2. The lack of mechanisms of subcontract optimisation taking into account non-renewable resources.
3. The inability to reschedule the works with the appearance of additional work. At the same time, the method provides the ability to specify the probabilistic duration and cost of the work.

We consider the application of the NPT, SGA and MGO- methods to the subcontracting scheduling problem within a project company. A detailed statement of the problem is given in [1]. The aim is to minimise the total subcontract volume.

The results of applying the NPT, SGA and MGO-methods using MS Project, RDO-studio and MEIS optimisation module respectively are shown in Fig. 3.

The following input information have been used: (1) 10 projects with 35 operations; (2) 10 employees; 3) time interval T = 430 days (1 year and 3 months); 4) time limit early and late start of the operations is determined by the shift in the provisional operations start dates for 2 weeks to the right or left along the time axis.

MS Project 2007 provides the opportunity for resource reallocation (with smoothing) in order to avoid exceeding the own renewable resources availability. The x-axis is the time intervals (each of which lasts 12 days); the y-axis is percentage utilisation of the company employees. The time intervals where the use of subcontract is necessary are shown in figure as dark stripes above the horizontal line at the 100% level.

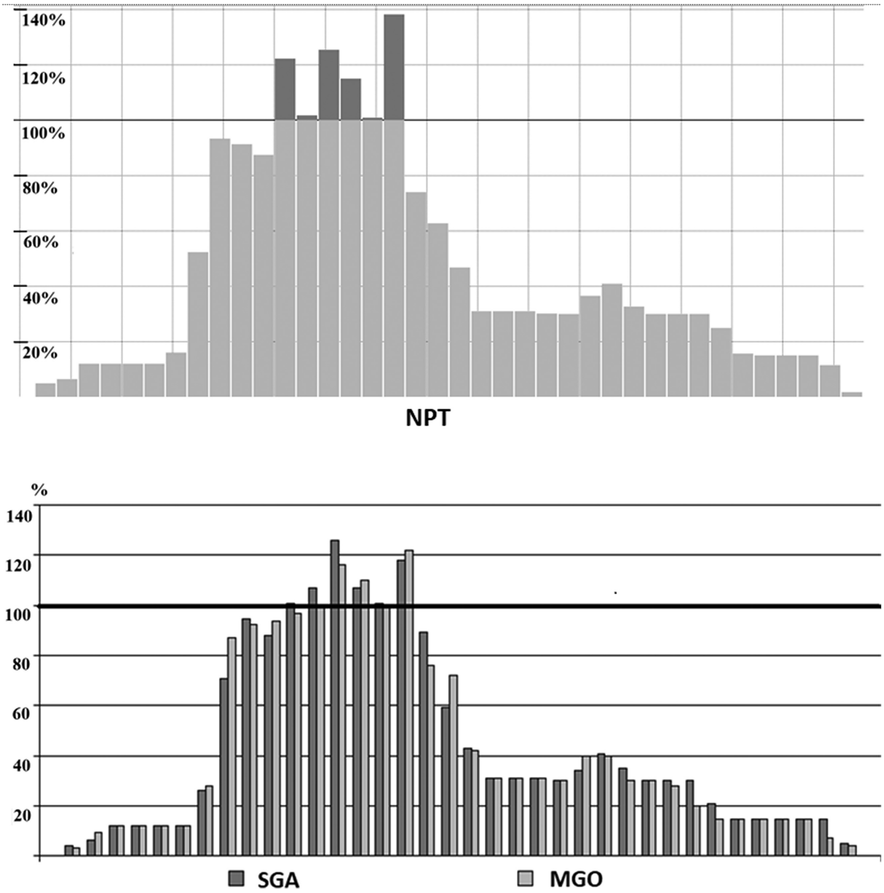


Fig. 3. Percentage utilisation of the company employees in the NPT, SGA and MGO-methods

As we can see, the resources deficit reaches 40% of the own resources volume when NPT-method applying, and reaches 25% when SGA and MGO-method applying. The MGO-method yields slightly better results than the SGA-method by applying a modified GA.

Thus, the MGO-method is the most preferable when subcontracting scheduling.

7 Conclusion and Future Work

In this paper, a comparative analysis in considering subcontracting scheduling problem has been conducted.

As a result of the comparative analysis, the advantages of the method of multiagent genetic optimisation in terms of solving the problem of subcontracting scheduling have been revealed. The multiagent genetic optimisation method takes into account the non-renewable resources, allows implementing different resource allocation strategies using simulation and multiagent modeling, takes into account the decision-makers behaviour using knowledge-based modeling, and allows optimising subcontract resources via analysis of alternative work schedules using genetic algorithms and simulation, reschedules the works using numerical methods of uncertainty removing and simulation.

The aim of future research is to extend and apply the developed MGO-method for the technological and logistics scheduling. The technological and logistics scheduling is complicated by consideration the production plan for the units of output and the availability of additional technological support operations, which are strictly related to the number of the completed basic technological operations on the industrial unit.

Acknowledgments. This work is supported by Act 211 Government of the Russian Federation, contract № 02.A03.21.0006.

References

1. Aksyonov, K., Antonova, A.: Application of simulation and intelligent agents to solve project management problem. Int. J. Comput. Sci. Eng. Inf. Technol. Res. **3**(1), 321–330 (2013)
2. Aksyonov, K., Antonova, A.: Multiagent genetic optimisation to solve the project scheduling problem under uncertainty. Int. J. Adv. Softw **7**(1&2), 1–19 (2014)
3. Aksyonov, K.A., Bykov, E.A., Aksyonova, O.P., Antonova, A.S.: Development of real-time simulation models: integration with enterprise information systems. In: The Ninth International Multi-Conference on Computing in the Global Information Technology, pp. 45–50 (2014)
4. Kureichik, V.M., Malioukov, S., Kureichik, V.V., Malioukov, A.: Genetic algorythms for applied CAD problems. Springer, Berlin (2009)
5. Moder, J., Elmaghraby, S.: Handbook of operations research: foundations and fundamentals, vol. 1, 2nd edn. Van Nostrand-Reinhold, New York (1978)
6. Pritsker, A., Happ, W.: GERT: graphical evaluation and review technique: Part I, fundamentals. J. Ind. Eng. **17**(6), 267–274 (1966)
7. Rzevski, G., Himoff, J., Skobelev, P.: MAGENTA technology: a family of multi-agent intelligent schedulers. In: International conference on multi-agent systems: Workshop on Software Agents in Information Systems and Industrial Applications 2 (SAISIA), Germany: Fraunhofer IITB (2006)
8. Vittikh, V., Skobelev, P.: Multiagent interaction models for constructing the needs-and-means networks in open systems. Autom. Remote Control **64**, 162–169 (2003)

Feature Based Opinion Mining for Restaurant Reviews

Nithin Y.R and Poornalatha G.(✉)

Department of Information and Communication Technology,
Manipal Institute of Technology, Manipal University,
Manipal 576104, Karnataka, India
yrnithin@gmail.com, poornalatha.g@manipal.edu

Abstract. Product reviews or customer feedback has become a platform for retailers to plan marketing strategy and also for new customers to select their appropriate product. Since the trend of e-commerce is increasing, an amount of customer reviews also has been increased to a greater extent. Consequently, it becomes a tough task for retailers as well as customers to read the reviews associated with the product. Sentiment analysis resolves this issue by scanning through free text reviews and providing the opinion summary. However, it does not provide detailed information, such as features on which the product is reviewed. Feature-based sentiment analysis methods increases the granularity of sentiment analysis by analyzing polarity associated with features in the given free text. The main objective of this work is to design a system that predicts polarity at aspect level and to design a score calculating scheme that defines the extent of polarity. Obtained feature - level scores are summarized according to users' priority of interest.

Keywords: Natural language processing · Aspects · Reviews · Free text · Star rating

1 Introduction

The web supports massive loads of user-generated information that describes customer opinions associated with commodities. These are considered to be valuable for consumers to make purchasing decisions and for business corporations to make marketing decisions. Product reviews are referred as an essential part of an online store's branding and marketing. They help build trust and loyalty. Customer reviews also describes the uniqueness of a particular product compared to other products. As per recent surveys [1], almost 70% of customers consult reviews or ratings before making a final purchase, 63% of consumers are more likely to purchase from a site if it has product ratings and reviews, 80% of consumers have changed their mind about purchases based on negative information they have found online. But manually scanning through the huge set of reviews to retrieve helpful decisions is complex and time-consuming. Ultimately

S.M. Thampi et al. (eds.), *Advances in Signal Processing and Intelligent Recognition Systems*,
Advances in Intelligent Systems and Computing 678, DOI 10.1007/978-3-319-67934-1_27

sentiment analysis and opinion mining have gained importance in the automatic analysis of user reviews and retrieving relevant information to users.

Opinion text classification at document or sentence level does not suffice real time applications as document level analysis focus on entity rather than its features. For exhaustive analysis, it is required to explore the occurrence of product features and determine the polarity on every product feature.

This work addresses the problem of summarizing opinions with respect to product features. It involves a syntactic and unsupervised method of finding polarity (positive, negative or neutral) from freetext. The process takes freetext as an input and uses NLP procedures to identify opinion words related to the specific aspect. Lexicon and dictionary-based approaches are used to calculate polarity associated with these opinion words. This result is represented using graphs and star rating schemes.

The organization of this paper is as follows. Section 2 explains the basics of sentiment analysis and various research contributions to the field of sentiment analysis. Section 3 reveals the methodology used in building the proposed system. This section includes details about the calculation of feature level sentiment scores and sentiment score summarization. Section 4 describes the datasets used by the proposed system. This section also includes the graphical analysis of opinion summarization. Section 5 gives the conclusion of the paper which is followed by references.

2 Literature Survey

Liu [2] defines the concept of opinion as a quintuple with the parameters like entity name (e(i)), aspect of entity (a(ij)), sentiment associated with aspect (s(ijkl)), holder of opinion (h(k)) and time at which opinion expressed (t(l)). These parameters are represented with indices to show that they correspond to each other.

The sentiment associated with entity can be represented in terms of polarity or expressed in terms of strength/intensity levels, e.g., 1 to 5 star rating is used in most of the review sites on the Web.

As described in [3], there are several research directions in sentiment analysis which includes feature-based opinion mining, comparison-based opinion mining, cross-domain sentiment analysis etc. The task of feature-based opinion mining involves an approach of discovering the entity objects on which opinions have been conveyed in sentence level and later polarity is determined on the component of the entity object. Entity objects are referred as the topic of interest which can be service, commodity, individual person, organization, event etc. and components are known as aspects, attributes or features. This type of fine-grained analysis is required in many practical applications because it helps to analyze which of the aspects relating to the product is liked or disliked by consumers. Comparison-based opinion mining focuses on finding sentiment on comparative opinions[1]. The research area of Cross-domain sentiment analysis is considered to be significant because classification of sentiment is tuned to a specific domain from which training data is retrieved and the same sentiment classifier may work inefficiently on test data from a different domain. The work [4] is considered to be useful in this regard.

Text classification with respect to sentiment analysis can be done at three different granularity levels [5], that is Document level, Sentence level and Entity - Aspect level. Document level sentiment classification determines whether given document expresses positive or negative sentiment by treating an entire document as a single unit. The task of Sentence level sentiment classification deals with determining polarity at the sentence level. Sentence sentiment classification is done on the assumption that each sentence gives single opinion given by the single author. This assumption may not be true in some situations where complex sentences appear. Entity - Aspect level sentiment analysis (also known as word level classification of sentiments) is precise compared to sentence and document level sentiment classification. It is based on the fact that an opinionated sentence consists of an opinion word and target word (entity/feature). For example in the sentence "Even Though the camera is no good, the phone works well", opinion on the phone as an entity is positive but feature as the camera has a negative opinion. It involves two tasks - feature extraction and feature sentiment classification. The task of feature extraction is based on finding opinionated expressions and its target object from the given sentence. After feature extraction, the task of feature sentiment classification determines the sentiment associated with each aspect. Early approaches were based on supervised learning approaches. But these techniques were proved to be inefficient while scaling massive collection of application domains. Dealing with sentiment shifters (negation words like not, neither, none etc.) and "but" clauses (situations wherein 2 sentences are joined by but conjunction) which alter the actual polarity of opinion words were considered as serious issues. Lexicon based approaches were designed to resolve these issues. The work of [6] uses a lexicon based approach. The flow of the approach used involves 4 steps, that is identifying sentiment words/phrases, processing sentiment shifters, processing "but" clauses and sentiment classification. In [3], Liu listed out some of the rules in BNF form to extract the feature from free text.

Processing (compiling) opinion words is considered to be a significant part of sentiment analysis. This can be done in three approaches: manual, dictionary-based and corpus based. Manual approach consumes more time compared to other two automated approaches. Dictionary-based approaches uses a list of synonyms and antonyms to compile given opinion words. Online dictionaries and frameworks like wordnet is used to accomplish this task. In [7], supervised learning was used to compile opinion words. The work [8] have utilized this approach for sentiment classification. Corpus based approaches uses opinion Lexicons to

[1] Comparative opinions involve comparison with other similar objects. For example, "Price of this phone is expensive" is an example for a regular opinion, while "price of this phone is better than phone-x" is a comparative opinion.

process opinion words. The work [9] used English lexicons in analyzing English customer reviews whereas the work [10] used Chinese lexicon in analyzing Chinese restaurant reviews.

3 Methodology

The system architecture of proposed design is depicted in Fig. 1.

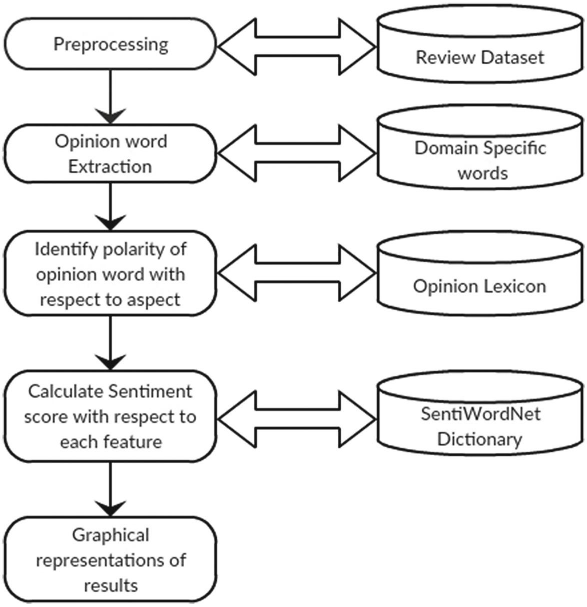

Preprocessing
Review Dataset
Opinion word Extraction
Domain Specific words
Identify polarity of opinion word with respect to aspect
Opinion Lexicon
Calculate Sentiment score with respect to each feature
SentiWordNet Dictionary
Graphical representations of results

Fig. 1. System architecture

3.1 Preprocessing

The input text is given to the system either by writing text in a textbox or by uploading a text file. The task of preprocessing begins with the splitting of text input in terms of sentences. The non - alphanumeric characters are removed from each of these sentences. This modified sentence is checked for the presence of domain-specific words. Sentences which does not possess any domain-specific words will be eliminated. There are chances of eliminating sentences containing pronouns that actually refers domain-specific words. These occurrences can be avoided by replacing pronouns with appropriate words which are actually being referred.

Pronoun Resolution: In this task of preprocessing, only simple pronouns are focused to be replaced. Replacing other types of pronouns may cause replacing of domain-specific words with other words. List of simple pronouns are taken from pronoun table specified in [11]. Here reflexive pronouns are not considered because replacing reflexive pronoun is ambiguous.

Algorithm 1. Pseudo Code For Pronoun Resolution

```
Input: Free text (Input_text)
Output: Text with pronouns replaced with words which is being referred
Result_text = Input_text
Annotate Input_text
for Each Input_coref_chain in Input_coref_chains do
    if Input_coref_chain_length > 1 then
        Annotate Result_text
        for Each mention in Result_coref_chain representing Input_coref_chain
         do
            if Result_coref_chain != null then
                if mention ∈ pronoun_list then
                    Replace mention by Representative_mention
            End for
    End for
Output Result_text
```

Algorithm 1 represents the Pseudo code used for pronoun resolution. This pseudo code can be implemented by using Stanford coreNLP [12] tools. It takes the free text as an input and gives an output text where pronouns are replaced by the word which is actually being referred. Annotation of the input text produces a group of coreference chains. Each of this chain has a representative mention[2] and a group of mentions[3]. Each of the mentions in coreference are checked for the presence in the pronoun list. If presence is confirmed mention will be replaced by the representative mention.

3.2 Opinion Word Extraction

In this process, the dependency relation between opinion word and domain-specific words (Aspect and aspect related words) is checked. A set of dependency rules is framed using Stanford parser tool [13]. Opinion words are extracted according to the list of rules specified in Table 1. Rules are framed using three attributes namely governor word, dependent word, and the relation between the governor and dependent words. Conventions used for framing rules are as follows:

[2] Representative mention - a special word in the sentence.

[3] Mentions are the words present in other sentences referring representative mention.

Table 1. Rules for extracting opinion words

Rule no	Relation	Governor word	Dependent word	Action
1	root	-	word	2(word), 11(word)
2	nsubj	word	aspect	ExtractOW(word), 3(word)
3	nsubj	word	aspect	4(word)
4	nsubj	word	aspect	5(word)
5	advmod	word	word1	ExtractOW(word1)
6	advcl	word	word1	ExtractOW(word1)
7	amod	word	word1	ExtractOW(word1)
8	cc	word	"but"	7(word)
9	nsubj	word	aspect	ExtractOW(word)
10	amod	aspect	word	ExtractOW(word)
11	rcmod	aspect	word	ExtractOW(word)
12	advmod	aspect	word	ExtractOW(word)
13	nsubj	word	aspect	12(word)
14	cop	word	word1	ExtractOW(word)
15	nsubj	word	aspect	14(word)
16	acomp	word	word	ExtractOW(word1)

- n(word) - Applying rule 'n' (n: rule number) to the given word.
- ExtractOW(word) - Extract word as opinion word.

The extracted opinion words are checked for negation by using dependency rules specified in Table 2. If extracted opinion word is marked with negation, then the polarity associated with opinion word changes (positive word changes to the negative word and vice versa).

Table 2. Rules for checking negation

Rule no	Relation	Governor word	Dependent word	Action
17	neg	word	-	Mark negative
18	conj_neg cc	-	word	Mark negative
19	Pobj	"not"	word	Mark negative

3.3 Identifying Polarity of Opinion Word and Calculating Sentiment Scores

An algorithm is designed to calculate sentiment score of the opinion word. This algorithm uses Opinion Lexicon [14] and SentiWordNet [7] to calculate sentiment score. Algorithm 2 represents the pseudo code for calculating sentiment scores. It takes opinion word as input and gives sentiment score with respect to M_{FS}[4]. Scores from SentiWordNet dictionary are obtained by using extract [15] method. Opinion Lexicon contains a list of positive and negative words. Some of these words are considered as trendy words which frequently appear in social media content. These words may not appear in SentiWordNet dictionary. Opinion Lexicon decides whether the word is positive, negative or neutral. In this score calculation, 0 is considered as minimum score and M_{FS} is considered as maximum score. The score $M_{FS}/2$ indicates neutral. If the given word is classified as positive by opinion lexicon, then the score is initialized as $(M_{FS}/2) + 1$. This indicates that the word is positive with respect to the given entity feature. If the word is classified as negative then the score is initialized as $(M_{FS}/2) - 1$, indicating the word is negative. The extent of polarity is determined by SentiWordNet scores. If the word is positive SentiWordNet score will be the positive number between 0 and 1. If it is negative the score will be a negative number 0 and -1. For opinion words marked as negative (using dependency rules), the calculated score is subtracted from maximum feature score.

Algorithm 2. Pseudo Code To Calculate Sentiment Score Associated With The Opinion Word

```
Input: word (Extracted opinion word)
Output: Sentiment score
score = 0
if word ∈ Positive_words_list then
    score = $(M_{FS}/2) + 1$
else if word ∈ Negative_words_list then
    score = $(M_{FS}/2) - 1$
else
    score = $M_{FS}/2$
    if SentiWordNet_score(word) != 0 then
        score = score + $(M_{FS}/2)$*SentiWordNet_score(word)
score = score + $((M_{FS}/2) - 1)$*SentiWordNet_score(word)
output score
```

3.4 Feature Level Summarization of Sentiment Scores

The proposed system summarizes feature level sentiment scores in 2 ways, by bar graph and by using star rates. All the sentiment scores calculated are grouped based on the specific aspect and an average is computed. This average will be the finalized feature level score.

[4] M_{FS} - Maximum feature score awarded to the entity feature with respect to the opinion words.

3.4.1 Bar Graph Summarization

In this type of summarization, all the finalized scores are represented using a bar graph. Google bar chart API [16] is used in this regard.

3.4.2 Star Rating

Under this type of summarization, a star rating based on finalized score is displayed. This star rating can be calculated in 4 ways. The calculation is done by taking maximum feature score as 10 and number of features as 5.

Case 1 - Simple Average: Average of all the five features can be represented in terms of stars (in 0–5 range) as follows:

$$SA_1 = \frac{FS_1 + FS_2 + FS_3 + FS_4 + FS_5}{10} \tag{1}$$

where,

- SA_c - Star Average for case ‘c’.
- FS_i - Feature scores for feature with ‘i’th priority.

Case 2 - Weighted Average: In this case, the star rating is calculated according to the priority. Feature with the highest priority gets the highest weight. Star rating in this scenario is calculated as follows:

$$SA_2 = \frac{FS_1 \times 5 + FS_2 \times 4 + FS_3 \times 3 + FS_4 \times 2 + FS_5 \times 1}{30} \tag{2}$$

Case 3 - Selected Feature [Type 1]: In this case few features are selected out of given 5 features. Here selected features are given weightage priority wise. Remaining features are given less but equal priority. For 5 features star rating is calculated in five ways:

For 1 Feature selection,

$$SA_3 = \frac{FS_1 \times 5 + FS_2 \times 2.5 + FS_3 \times 2.5 + FS_4 \times 2.5 + FS_5 \times 2.5}{30} \tag{3}$$

For 2 Feature selection,

$$SA_3 = \frac{FS_1 \times 5 + FS_2 \times 4 + FS_3 \times 2 + FS_4 \times 2 + FS_5 \times 2}{30} \tag{4}$$

For 3 Feature selection,

$$SA_3 = \frac{FS_1 \times 5 + FS_2 \times 4 + FS_3 \times 3 + FS_4 \times 1.5 + FS_5 \times 1.5}{30} \tag{5}$$

For 4 and 5 Feature selection,

$$SA_3 = \frac{FS_1 \times 5 + FS_2 \times 4 + FS_3 \times 3 + FS_4 \times 2 + FS_5 \times 1}{30} \tag{6}$$

Case 4 - Selected Feature [Type 2]: This case is similar to the case 3 but only selected features are taken into considerations and weights are assigned priority wise. Rest of the features are not given any weightage. Star rating calculation is this scenario can be done as follows:

For 1 Feature selection,

$$SA_4 = \frac{FS_1 \times 5}{10} \tag{7}$$

For 2 Feature selection,

$$SA_4 = \frac{FS_1 \times 5 + FS_2 \times 4}{18} \tag{8}$$

For 3 Feature selection,

$$SA_4 = \frac{FS_1 \times 5 + FS_2 \times 4 + FS_3 \times 3}{24} \tag{9}$$

For 4 Feature selection,

$$SA_4 = \frac{FS_1 \times 5 + FS_2 \times 4 + FS_3 \times 3 + FS_4 \times 2}{28} \tag{10}$$

For 5 Feature selection,

$$SA_4 = \frac{FS_1 \times 5 + FS_2 \times 4 + FS_3 \times 3 + FS_4 \times 2 + FS_5 \times 1}{30} \tag{11}$$

Star-rating schemes specified above enables users to visualise feature level sentiment scores. This visualisation of scores is done according to user's priority of interest.

- Case-1 star-rating visualizes the feature-level summary by giving equal priorities to all features.
- Case-2 star-rating visualizes the feature-level summary according to user's priority of interest. Here priority is given to all features.
- Case-3 star-rating also visualizes the feature-level summary with respect to user's priority of interest. Here priority is given to a selected set of features. Rest of non-selected features are given equal but less priorities.
- Case-4 is a priority-based visualization of star-rating. Here priority is only given to a set of features.

4 Results

The proposed system is designed to work on restaurant reviews. Reviews that belong to Yelp Dataset Challenge [17] are taken as input. This dataset contains reviews on 1,44,072 enterprises, represented with unique business IDs. Out of these 48,485 business IDs belongs to restaurant category. Of these restaurant business IDs, reviews of restaurants with business IDs 4P-vTvE6cncJyUyLh73pxw and 4uiijOUDzc-DeIb2XcK_A were used. Restaurant with business ID 4P-vTvE6cncJyU- yLh73pxw has 23 reviews and latter one has 37 reviews. Each of these reviews is stored in a separate text file. Sentiment score calculation is done on 5 restaurant aspects - ambience, service, staff, food and price.

Each of these reviews is given as input to the system in terms of text files, through a user interface as shown in Fig. 2(A).

Feature priority for star calculation is set through the same user interface, using selection boxes. Type 1 selection is used for selecting priorities for case 2 star rating calculation and type 2 selection is used for selecting features for case 3 and case 4 star rating calculation.

In this instance, aspect priorities are set in the order ambience, service, staff, food and price for Type-1 selection. Type-2 selection is done by selecting three of five aspects and aspect priorities are set in the order food, service and ambience.

Each of these priority selection and resulting casewise star rating explains a specific scenario which is explained as follows:

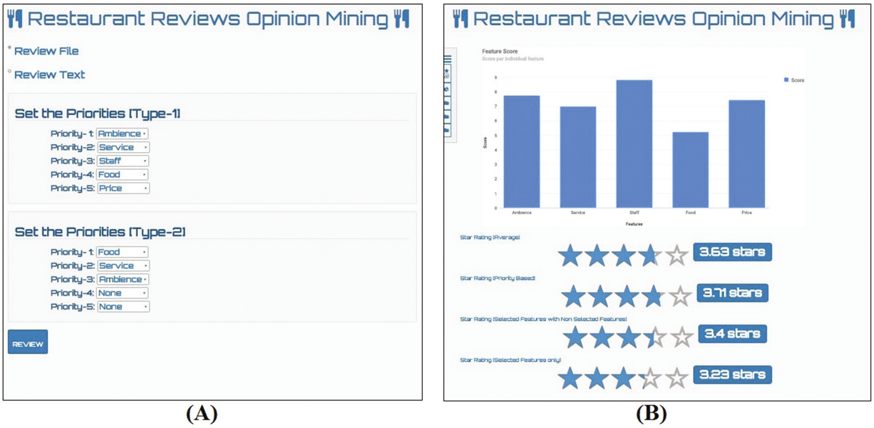
- Case-1 explains the scenario wherein user is interested in all the features of restaurant.
- Case-2 explains the scenario where the user wants to know the restaurant reviews summary and priority of interest is given for all five features of a restaurant in the order ambiance, service, staff, food and price.
- Case-3 explains the scenario where the user is more interested in features food, service and ambiance. Other features like staff and price are given less priority. Here all the features are taken into consideration but selected features are given more priority.
- Case-4 explains the scenario where user is only interested in features food, service and ambiance. Rest of the features are not given any importance.

After the submission of input details, the system receives the text input. Irrelevant symbols and sentences will be removed from this input text and opinion words will be extracted by using dependency rules. Extracted opinion words undergo the process of score calculation and finalized aspect scores will be calculated. Finalized scores are summarized as shown in Fig. 2(B). Finalized aspect scores are represented in bar graphs and following 4 star rating representations depict case-1, case-2, case-3 and case-4 respectively.

Each of these extracted opinions and their respective scores is taken as instances. Classification of these opinions can be done in 3 labels, i.e. positive, negative and neutral. The comparison of classified label and actual label is represented with the notation specified in Table 3. A confusion matrix is obtained as shown in Table 4.

Accuracy for this confusion matrix will be the fraction involving the sum of occurrences of labels that appear in matrix diagonal and sum of occurrence of all labels.

The proposed system is designed with 2 approaches. Approach-1 involves lesser domain-specific words and dependency rules compared to domain-specific words and dependency rules used in approach-2. Approach-1 uses Stanford



Restaurant Reviews Opinion Mining

○ Review File

○ Review Text

Set the Priorities [Type-1]

Priority-1: Ambience

Priority-2: Service

Priority-3: Staff

Priority-4: Food

Priority-5: Price

Set the Priorities [Type-2]

Priority-1: Food

Priority-2: Service

Priority-3: Ambience

Priority-4: None

Priority-5: None

REVIEW

Restaurant Reviews Opinion Mining

Star Rating [Average] 3.63 stars

Star Rating [Priority Based] 3.71 stars

Star Rating [Selected Features with Non Selected Features] 3.4 stars

Star Rating [Selected Features only] 3.23 stars

Fig. 2. (A) User interface to enter text input and set priorities. (B) Aspect level opinion summary.

Table 3. Notations used for label comparison

Label short name	Label description
T_{Pos}	True positive
$FNeg_{Pos}$	False negative, actual positive
$FNeu_{Pos}$	False neutral, actual positive
$FPos_{Neg}$	False Positive, actual negative
T_{Neg}	True negative
$FNeu_{Neg}$	False neutral, actual negative
$FPos_{Neu}$	False positive, actual neutral
$FNeg_{Neu}$	False negative, actual neutral
T_{Neu}	True neutral

Table 4. Confusion matrix

	Classified positive	Classified negative	Classified neutral
Actual positive	T_{Pos}	$FNeg_{Pos}$	$FNeu_{Pos}$
Actual negative	$FPos_{Neg}$	T_{Neg}	$FNeu_{Neg}$
Actual neutral	$FPos_{Neu}$	$FNeg_{Neu}$	T_{Neu}

Table 5. Accuracy details

Business ID	Approach-1	Approach-2
4P-vTvE6cncJyUyLh73pxw	72%	78%
4uiijOUDzc-DeIb2XcK_A	64%	71%

CoreNLP tool to split sentences whereas Approach-2 uses Apache OpenNLP [18] tool in sentence splitting. Reviews related to the business IDs 4P-vTvE6cncJyUyLh73pxw and 4uiijOUDzc-DeIb2XcK_A are taken as input to the system. Accuracy details of approaches used are given by the Table 5.

In some of the instances, it is being observed that Opinion Lexicons work perfect but SentiWordNet scores appear to be inappropriate. For Example, the word expensive is classified as negative with respect to aspect price by opinion lexicon, but the method of extracting SentiWordNet score gives a positive score. This alters score calculation and hence the polarity associated with opinion word.

An analysis on star rating is performed by setting priority boxes in a specific order. This priority-selection order is maintained to be the same for all review inputs. For case 2, aspect priorities are given in the following order - ambiance, service, staff, food and price. For case 3 and 4, three of five aspects are set in the following order - food, service and ambiance. A graph has been plotted by taking star-rating along y-axis and review inputs along x-axis. Two groups of review input files, each belonging to the business IDs 4P-vTvE6cncJyUyLh73pxw and 4uiijOUDzc-DeIb2XcK_A are taken as input for star-analysis. Each of these groups contains 10 review input files. Analysis graphs of review inputs related to business ID 4P-vTvE6cncJyUyLh73pxw are shown in figures Fig. 3(A) and (B). Analysis graphs related to the business ID 4uiijOUDzc-DeIb2XcK_A is represented in figures Fig. 4(A) and (B).

Among the line-graphs represented by figures Figs. 3(A), (B) and 4(A), (B), it has been observed that lines representing user star rating differ in its variations with respect to other lines. The reason behind this behavior is that star-rating given by the user is an overall star-rating whereas star-rating calculated in four different cases is based on specific set of features. For each Review input file, corresponding case-wise star-rating results are obtained. Thus these analysis graphs explain how priority wise opinions differ from one review to other.

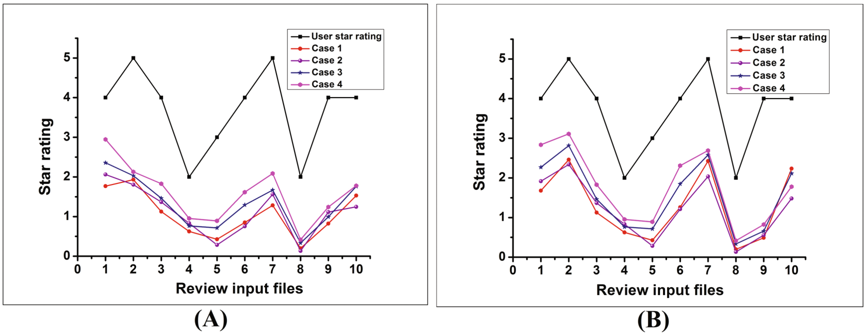
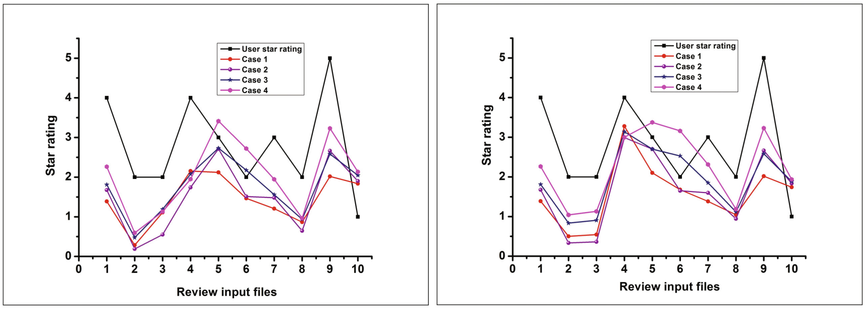


Fig. 3. Star - rating analysis of 4P-vTvE6cncJyUyLh73pxw review inputs under Approach-1 (A) and Approach-2 (B)



(A) (B)

Fig. 4. Star - rating analysis of 4uiijOUDzc-DeIb2XcK_A review inputs under Approach-1 (A) and Approach-2 (B)

Lines representing case-1 follows the star-average Eq. (1) and lines representing case-2 follows the star-average Eq. (2). Out of five features, three features have been selected. Consequently lines representing case-3 follows the star-average Eq. (5) and lines representing case-4 follows the star-average Eq. (9).

5 Conclusions

This research work aims at designing a system that identifies opinion polarities hidden in the given free text at feature-level. Calculation of feature-level sentiment score represents the extent of polarity. Dependency rules used in the system were found to be beneficial in finding opinion words that determine polarity. Opinion lexicons and frameworks like SentiWordNet were helpful in determining sentiment scores. Visual representation of calculated sentiment scores is done using bar graph and star rating calculation schemes. The accuracy of the system can be increased by incorporating more dependency rules and by improving the method of extracting SentiWordNet scores so that score calculation scheme becomes more domain specific. The addition of dependency rules increases the efficiency of the opinion word extraction process. Resolving the problem of word sense disambiguation makes the system more domain specific.

References

1. Why online store owners should embrace online reviews. https://www.shopify.in/blog/15359677-why-online-store-owners-should-embrace-online-reviews. Accessed 21 Apr 2017
2. Liu, B.: Sentiment analysis and subjectivity. In: Handbook of Natural Language Processing, pp. 627–666 (2010)
3. Liu, B.: Sentiment analysis and opinion mining. In: Synthesis Lectures on Human Language Technologies, pp. 1–167 (2012)

4. Bagheri, A., Saraee, M., de Jong, F.: Care more about customers: Unsupervised domain-independent aspect detection for sentiment analysis of customer reviews. Knowl.-Based Syst. **52**, 201–213 (2013)
5. Petz, G., Karpowicz, M., Fürschuß, H., Auinger, A., Stříteský, V., Holzinger, A.: Computational approaches for mining user's opinions on the web 2.0. Inf. Process. Manage. **50**(6), 899–908 (2014)
6. Ding, X., Liu, B., Yu, P.S.: A holistic lexicon-based approach to opinion mining. In: Proceedings of the 2008 International Conference on Web Search and Data Mining, pp. 231–240 (2008)
7. Baccianella, S., Esuli, A., Sebastiani, F.: Sentiwordnet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining. In: Proceedings of LREC, pp. 2200–2204 (2010)
8. Dongre, A.G., Dharurkar, S., Nagarkar, S., Shukla, R., Pandita, V.: A survey on aspect based opinion mining from product reviews. Int. J. Innovat. Res. Sci. Eng. Technol. **5**(2), 1415–1418 (2016)
9. Hu, M., Liu, B.: Mining and summarizing customer reviews. In: Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 168–177 (2004)
10. Zhu, J., Wang, H., Zhu, M., Tsou, B.K., Ma, M.: Aspect-based opinion polling from customer reviews. IEEE Trans. Affect. Comput. **2**(1), 37–49 (2011)
11. Pronouns chart. http://www.grammarbank.com/pronouns-chart.html. Accessed 22 May 2017
12. Manning, C.D., Surdeanu, M., Bauer, J., Finkel, J., Bethard, S.J., McClosky, D.: The Stanford CoreNLP natural language processing toolkit. In: Association for Computational Linguistics (ACL) System Demonstrations, pp. 55–60 (2014)
13. De Marneffe, M.-C., Manning, C.D.: Stanford typed dependencies manual (2008)
14. Opinion lexicon. https://www.cs.uic.edu/~liub/FBS/sentiment-analysis.html. Accessed 22 May 2017
15. Sentiwordnet sample code. http://sentiwordnet.isti.cnr.it. Accessed 22 May 2017
16. Google bar chart. https://developers.google.com/chart/interactive/docs/gallery/barchart. Accessed 28 May 2017
17. Yelp dataset challenge. https://www.yelp.com/dataset_challenge. Accessed 28 May 2017
18. Apache opennlp developer documentation. https://opennlp.apache.org/docs/1.8.0/manual/opennlp.html. Accessed 28 May 2017

Exploring the Significance of Low Frequency Regions in Electroglottographic Signals for Emotion Recognition

S.G. Ajay(✉), D. Pravena, D. Govind, and D. Pradeep

Centre for Computational Engineering and Networking (CEN),
Amrita School of Engineering, Amrita Vishwa Vidyapeetham,
Coimbatore 641112, Tamilnadu, India
ajay190694@gmail.com, d.pravena@gmail.com, d_govind@cb.amrita.edu,
getpradeepd@gmail.com
http://www.amrita.edu/campus/coimbatore

Abstract. Electroglottographic (EGG) signals are acquired directly from the glottis. Hence EGG signals effectively represent the excitation source part of the human speech production system. Compared to speech signals, EGG signals are smooth and carry perceptually relevant emotional information. The work presented in this paper includes a sequence of experiments conducted on the emotion recognition system developed by the Gaussian Mixture Modeling (GMM) of perceptually motivated Mel Frequency Cepstral Coefficients (MFCC) features extracted from the EGG. The conclusions drawn from these experiments are two folds. (1) The 13 static MFCC features showed improved emotion recognition performance than 39 MFCC features with dynamic coefficients (by adding Δ and Δ Δ). (2) Low frequency regions in the EGG are emphasized by increasing the number of Mel filters for MFCC computation found to improve the performance of emotion recognition for EGG. These experimental results are verified on the EGG data available in the classic German emotional speech database (EmoDb) for four emotions such as (Anger, Happy, Boredom and Fear) apart from Neutral signals.

Keywords: EGG · MFCC · GMM · HTK · openEAR

1 Introduction

With the advancements in Machine Learning and Artificial Intelligence, the need for effective Human-Machine interaction has gained significant importance. The impact of emotional speech in Human-Machine interaction is less significant due to the fact that, machines cannot understand human's emotional state [13]. This has increased the need for analysis of emotions during the Human-Machine interaction. Usually, emotions of human beings are analyzed from the recorded speech signals. Rather than using recorded speech signals, Electroglottographic

S.M. Thampi et al. (eds.), *Advances in Signal Processing and Intelligent Recognition Systems*, Advances in Intelligent Systems and Computing 678,
DOI 10.1007/978-3-319-67934-1_28

(EGG) signals can be used for recognizing the emotions. Other than EGG signals, another approximation representing the glottal information (excitation source) is by using linear prediction residual signals [2,5], which can be derived from speech signals. By the informal listening to EGG signals, humans can identify the emotion it carries. With the availability of EGG data in different emotions in the classic German emotional speech database (EmoDb), the present work focuses on using EGG signals for emotion recognition.

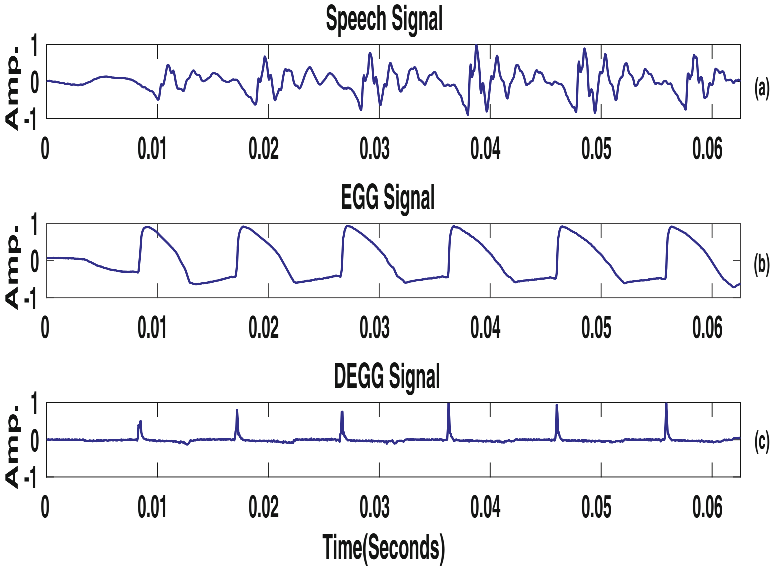


Fig. 1. Smooth nature of EGG signal. Plot (a) Speech signal, (b) EGG signal, (c) Differenced EGG signal.

Figure 1 clearly depicts the difference between the speech and the EGG signals. As seen from the Fig. 1, EGG signals are smooth compared to speech signals and also when they are differentiated, a sequence of impulses are produced which represents the glottal closure instants (GCIs). This indicates that the glottal information is clearly present in it and also it carries perceptually relevant emotional information (excitation source information) in the low frequency regions. This phenomenon is shown clearly in the Fig. 2. The analysis of this phenomenon is performed by calculating wide-band spectrogram for different emotional (Anger, Happy, Boredom, and Fear) utterances apart from Neutral signals present in the EGG data. The same vowel is chosen for representing the variations in different emotions through the spectrogram. Figure 2(f)–(j) shows the spectrogram of the EGG signal produced when the vowel /a/ is elicitated with the different emotional state of the same speaker. The corresponding EGG signal is plotted in Fig. 2(a)–(e). In all the spectrograms, the low frequency regions are very dark when compared to the high frequency regions, which shows that the

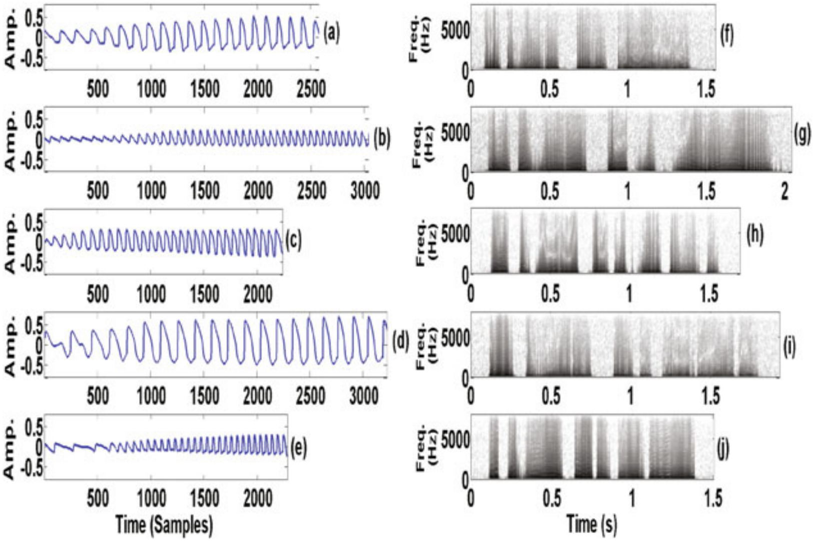


Fig. 2. Spectrogram analysis of glottal signals for the same utterance with different emotions. Plot (a)–(e) shows the glottal signals for Neutral, Anger, Happy, Boredom and Fear emotions respectively. Plot (f)–(j) are the corresponding wide-band spectrogram of glottal signals in (a)–(e).

availability of emotional information in glottal signals are more concentrated in the low frequency regions.

In this work, the state of the art perceptually motivated Mel Frequency Cepstral Coefficients (MFCC) are considered as features of the EGG signals for Gaussian Mixture Modelling [10]. Since the human perception of sound is in Mel scale, Mel filters were used for the computation of MFCC features in speech signals. MFCC's are widely used as features for many applications like Speaker Recognition, Speaker Identification, Speech Recognition, Emotion Recognition etc. [9,14,15]. The works presented so far in the literature shows that, emotion recognition is performed exclusively for emotive speech signals [1,7,11,16]. Pati et al. [12] used Residual MFCC (RMFCC) features from the linear prediction residual signals which is an approximation of glottal signals derived from the speech. Since the EGG signals have only glottal information, the proposed emotion recognition system uses MFCC features from it. So the work presented in this paper, attempts to experiment on the EGG signals by extracting the 13 static MFCC and the 39 dynamic MFCC features, by varying the number of filters in the Mel filter bank in order to emphasize the low frequency regions for computing MFCC. The organization of the work is as follows: Sect. 2 refers to the Development of emotion recognition system using EGG. Section 3 explains the Performance analysis of Emotion recognition using EGG signals. Summary and Conclusion of the present work are discussed in Sect. 4.

2 Development of Emotion Recognition System Using EGG

2.1 Production of EGG

EGG encompasses more emotional information which is captured at the time of elicitation. It is recorded through a device named Electroglottograph as shown in Fig. 3, which contains a pair of electrodes placed near the glottis region to capture the vocal fold vibrations during the production of speech [8]. It measures the vibration by passing a small amount of current between the contact area of the vocal folds. The impedance across the electrodes varies with respect to the vibrations of the vocal folds [6]. This variation of impedance produces quasi-periodic (non-stationary) signals known as the EGG signals.

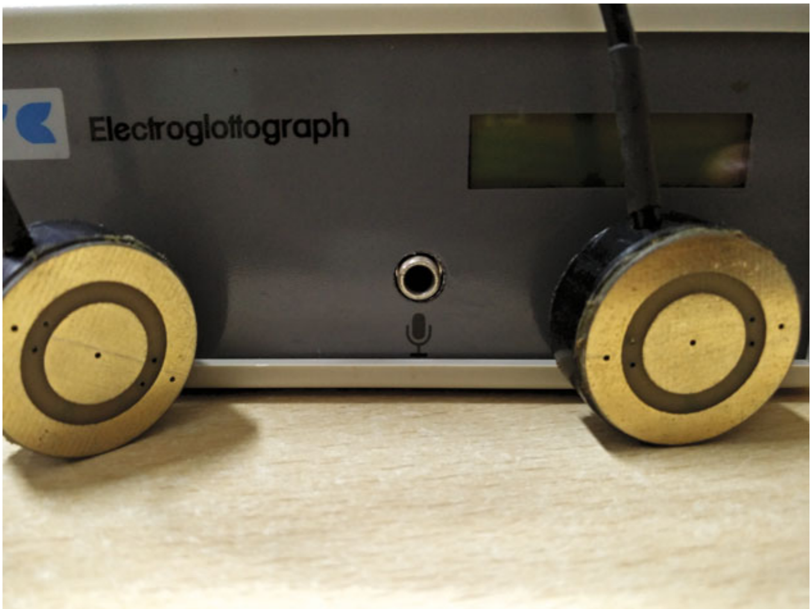


Fig. 3. Electroglottograph.

MFCC Feature Extraction from EGG. A nonlinear triangular Mel scale filter bank [14] as shown in Fig. 4 (filters are linearly placed in the low frequency regions (<1000 Hz) and logarithmically placed in the high frequency regions (>1000 Hz)) has the potentiality to emphasize the lower frequency components over the higher ones. Mel filters are designed to mimic human auditory perception of sound by concentrating more on the low frequency regions. As EGG signals are low frequency in nature, Mel Frequency Cepstral Coefficients can act as good features representing the emotional information present in the low frequency regions. In order to extract features, the non-stationary signal is divided into a smaller number of stationary frames of size 20 ms using Hamming window. Hamming window is used to avoid spectral leakage. A Hamming window shift of 10 ms is used. Along with the 13 static MFCC (velocity features), Δ and $\Delta\ \Delta$ (acceleration features) are extracted from each frame and they are combined with the 13 MFCC features to make the feature vector dimension as 39. By increasing the filter banks ranging from $(14, 16, 18....46)$, 13 and 39 MFCC features are extracted individually.

3 Performance Analysis of Emotion Recognition Using EGG Signals

The performance of emotion recognition using EGG signals is analyzed in the classic German emotional speech database (EmoDb) [3] which includes a simultaneous recording of Speech and EGG signals. The database was developed for six different emotions Anger, Happy, Fear, Boredom, Sad, Disgust apart from Neutral signals with 10 professional actors (5 Male and 5 Female) using 10 neutral sentences spoken in six emotions. Out of six emotions, four emotions (Anger, Happy, Fear, Boredom) along with Neutral signals are considered for this emotion recognition analysis. Each speech sample is recorded at a sampling rate of 48 KHz with 16 bits per sample resolution. In this work Speech and EGG signals of German emotional speech data are separated, and the separated EGG signals are downsampled to 16 KHz. Training and Testing for the analysis are performed with 590 utterances. Out of 590 utterances, 474 utterances were taken for training the GMM's and 116 utterances were used for testing the GMM's. A series of experiments were conducted with the 13 static MFCC features and the 39 dynamic MFCC features with different filter bank coefficients. These cepstral features are trained with 512 Gaussian Mixture components as the training data is small and the trained GMM's are tested for the classification accuracy in different emotional classes. For implementing MFCC-GMM based emotion recognition system, we have used HTK (Hidden Markov Model) toolkit [17] in our experiments. The configuration files are given in the following link https://drive.google.com/drive/folders/0BzHkgLdbz2n-OGR5dnJHTXhwVTg?usp=sharing. From the experiments conducted the observations inferred are two folds.

EGG signals show better performance in classifying the emotions with the 13 static MFCC features than the 39 dynamic MFCC features (by adding Δ and $\Delta\ \Delta$) for the conventional Mel filter bank of size 28 as seen from Table 1.

The rationality in this performance is due to the fact that, while taking the 39 dynamic MFCC features the change in the dynamics of the vocal tract across different frames of an audio signal is accounted. Unlike speech signals, EGG signals contain only glottal information (excitation source information) and it is clearly captured by the 13 static MFCC features. Since EGG signals lacks the vocal tract information, accounting the dynamic features (Δ and $\Delta\ \Delta$) which

Table 1. Classification Accuracies(%) for emotion recognition in EGG from German EmoDb with the 13 and 39 MFCC features for the conventional filter bank of size 28.

Number of Gaussians	Accuracy(%)	
	Filter bank of size 28	
	13 MFCC	39 MFCC
8	49.14	54.31
16	64.66	56.03
32	60.34	58.62
64	71.55	63.79
128	76.72	67.24
256	75.86	67.24
512	77.59	68.10

represents the same is not helping in improving the recognition performance. Therefore the proposed work experiments only on the use of the 13 static MFCC features by increasing the number of filters in the low frequency regions, thereby giving more emphasis.

Tables 2 and 3 discusses the series of experiments conducted with the 13 static MFCC features by varying the number of Mel filters in the Mel filter bank.

Table 2. Classification Accuracies(%) for emotion recognition in EGG from German EmoDb with the 13 static MFCC features containing different filter bank coefficients.

Number of Gaussians	Accuracy(%)										
	Different sizes of Mel filter banks										
	14	**16**	**18**	20	22	24	26	28	30	32	34
8	54.31	52.59	50.00	50.00	50.86	52.59	49.14	49.14	51.72	51.72	52.59
16	59.48	56.03	62.93	59.48	62.93	58.62	63.79	64.66	61.21	64.66	62.93
32	65.52	68.10	67.24	67.24	65.52	66.38	68.10	60.34	65.52	65.52	66.38
64	70.69	67.24	70.69	68.97	72.41	71.55	70.69	71.55	71.55	70.69	69.83
128	74.14	72.41	77.59	73.28	71.55	75.86	71.55	76.72	75.86	75.86	74.14
256	**72.41**	**72.41**	**74.14**	76.72	72.41	77.59	72.41	75.86	78.45	76.72	72.41
512	69.83	73.28	75.00	76.72	73.28	76.72	76.72	77.59	79.31	77.59	75.86

It is evident from Tables 2 and 3 that, performance of the MFCC-GMM based emotion recognition system increases by increasing the number of Mel filters in the low frequency regions. This is because, EGG signals are low frequency in nature and therefore by keeping more filters in the low frequency regions, more emotional information is captured. The optimal performance with 80.17% is obtained while using 256 Gaussian mixtures with the 13 static MFCC features for the higher order filter bank of size 38. Also when the filters are increased

Table 3. Classification Accuracies(%) for emotion recognition in EGG from German EmoDb with the 13 static MFCC features containing different filter bank coefficients.

Number of Gaussians	Accuracy(%)					
	Different sizes of Mel filter banks					
	36	**38**	**40**	**42**	44	46
8	51.72	50.86	49.14	51.72	46.55	43.97
16	63.79	60.34	62.93	62.07	62.93	61.21
32	61.21	61.21	66.38	66.38	59.48	60.34
64	69.83	73.28	70.69	71.55	73.28	72.41
128	73.28	76.72	75.86	75.86	77.59	75.00
256	76.72	**80.17**	**79.31**	**79.31**	75.86	75.00
512	75.00	77.59	77.59	77.59	76.72	79.31

beyond 38, the recognition performance seem to degrade, this is due to the fact that, when filters are more denser at the low frequency regions, the width of the traingular filters decreases, this, in turn, fails to capture the relevant emotional information. Figure 4 shows the traingular Mel filter bank of size 28 and 36. It is evident from the Fig. 4 that, when filters are denser in the lower frequency regions more emphasis is given.

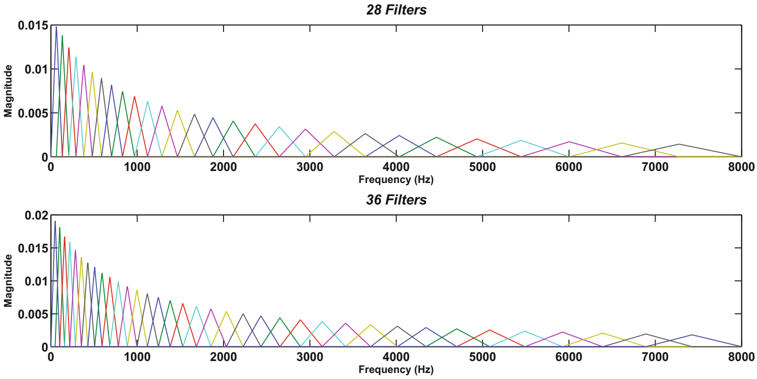


Fig. 4. Mel Filter banks of size 28 and 36.

4 Summary and Conclusion

The work proposed in this paper focuses on using the EGG signals for emotion recognition. As EGG signals are low frequency in nature and it approximates the glottal information during the production of the emotive speech signals, the perceptually motivated Mel Frequency Cepstral Coefficients (MFCC) are extracted from the same for Gaussian Mixture Modeling. The conclusions drawn from this work is as follows,

- The MFCC-GMM system with the 39 dynamic MFCC features with Δ and Δ Δ does not contribute for improved emotion recognition performance in case of the EGG signals, whereas the 13 static MFCC features give better performance for the same.
- In order to emphasize the low frequency components, increasing the number of Mel filters in Mel filter bank to a certain level for the computation of MFCC features helps to improve the emotion recognition performance.

The future work concentrates on building an emotion recognition system using the acoustic features derived from the Munich's openEAR toolkit [4] for the same EGG signals present in the classic German emotional speech database (EmoDb). The results obtained for the emotion recognition from the conventional state of the art MFCC-GMM can be verified or improved using other classification algorithms in Deep Networks.

References

1. Albornoz, E.M., Milone, D.H., Rufiner, H.L.: Spoken emotion recognition using hierarchical classifiers. Comput. Speech Lang. **25**, 556–570 (2011)
2. Ananthapadmanabha, T.V., Yegnanarayana, B.: Epoch extraction from linear prediction residual for identification of closed glottis interval. IEEE Trans. Acoust. Speech Sig. Process. **27**(4), 309–319 (1979)
3. Burkhardt, F., Paeschke, A., Rolfes, M., Sendlemeier, W., Weiss, B.: A database of German emotional speech. In: Proceedings of INTERSPEECH, pp. 1517–1520 (2005)
4. Eyben, F., Wöllmer, M., Schuller, B.: Opensmile: the Munich versatile and fast open-source audio feature extractor, pp. 1459–1462 (2010)
5. Govind, D., Prasanna, S.R.M.: Expressive speech synthesis: a review. Int. J. Speech Technol. **16**(2), 237–260 (2013)
6. Henrich, N., DAlessandro, C., Doval, B., Castellengo, M.: On the use of the derivative of electroglottographic signals for characterization of nonpathological phonation. J. Acoust. Soc. Am. **115**(3), 1321–32 (2004)
7. Kandali, A.B., Routray, A., Basu, T.K.: Emotion recognition from Assamese speeches using MFCC features and GMM classifier. In: IEEE Region 10 Conference (2008)
8. Kitzing, P.: Clinical applications of electroglottography. J. Voice **4**(3), 238–249 (1990)
9. Koolagudi, S.G., Rao, K.S.: Two stage emotion recognition based on speaking rate. Int. J. Speech Technol. **14**, 35–48 (2011)
10. Koolagudi, S.G., Rao, K.S.: Emotion recognition from speech using source, system, and prosodic features. Int. J. Speech Technol. **15**, 265–289 (2012)
11. Neiberg, D., Elenius, K., Laskowski, K.: Emotion recognition in spontaneous speech using GMMS. In: INTERSPEECH (2006)

12. Pati, D., Prasanna, S.R.M.: Processing of linear prediction residual in spectral and cepstral domains for speaker information. Int. J. Speech Technol. **18**(3), 333–350 (2015)
13. Prasanna, S.R.M., Govind, D.: Analysis of excitation source information in emotional speech. In: Proceedings INTERSPEECH, pp. 781–784 (2010)
14. Pravena, D., Nandhakumar, S., Govind, D.: Significance of natural elicitation in developing simulated full blown speech emotion databases, pp. 261–265 (2016)
15. Raviram, P., Umarani, S.D., Wahidabanu, R.S.D.: Isolated word recognition using enhanced MFCC and IIFS. In: Proceedings of the International Conference on Frontiers of Intelligent Computing: Theory and Applications (FICTA), vol. 199, pp. 273–283. Springer (2013)
16. Vondra, M., Vch, R.: Recognition of emotions in German speech using Gaussian mixture models. Multimodal Sig. **5398**, 256–263 (2009)
17. Young, S.J., Young, S.: The HTK hidden Markov model toolkit: design and philosophy (1993)

Tamil Speech Emotion Recognition Using Deep Belief Network(DBN)

M. Srikanth(✉), D. Pravena, and D. Govind

Centre for Computational Engineering and Networking (CEN),
Amrita School of Engineering, Amrita Vishwa Vidyapeetham,
Coimbatore 641112, Tamilnadu, India
srikanthmurali3@gmail.com, d.pravena@gmail.com, d_govind@cb.amrita.edu
http://www.amrita.edu/campus/coimbatore

Abstract. The proposed system shows the effectiveness of Deep Belief Network(DBN) over Gaussian Mixture model(GMM). The development of the proposed GMM-DBN system is by modeling GMM for each emotion independently using the extracted Mel frequency Cepstral Coefficient(MFCC) features from speech. The minimum distance between the distribution of features for each utterance with respect to each emotion model is derived as Bag of acoustic features(BoF) and plotted as histogram. In histogram, the count represents the number of feature distributions that are close to each emotion model. The BoF is passed in to DBN for developing train models. The effectiveness of the emotion recognition using DBN is empirically observed by increasing the Restricted Boltzmann machine(RBM) layers and further by tuning available parameters. The motivation is by testing the Classical German Speech emotion database(EmodB) with the proposed GMM-DBN system which gives the performance rate increase by 5% than the conventional MFCC-GMM system by empirical observation. Further testing of the proposed system over the recently developed simulated speech emotion database for Tamil language gives a comparable result for the emotion recognition. The effectiveness of the proposed model is empirically observed in EmodB.

1 Introduction

The Speech Emotion recognition (SER) can bridge gap between the Human-Machine interaction. The significant empirical observations are made decades ago [1] provides information about the importance of vocal cues for the expression of speech. Information about the emotional status of the speaker can improve the communication between the listeners and derive more understanding, especially the exact meaning in between words. The identification of emotion state in speech is termed as the SER. The SER is practically applied in many growing areas like automobile industry, robotics, e-learning centers, etc. The lack of understanding of emotion intelligence due to the unavailability of actual emotion data and procedures to collect data are existing due to previous survey works [2].

S.M. Thampi et al. (eds.), *Advances in Signal Processing and Intelligent Recognition Systems*, Advances in Intelligent Systems and Computing 678,
DOI 10.1007/978-3-319-67934-1_29

The difficulty is present in understanding emotion at the levels of physiology and psychology is present even before going to the actual analysis of speech data. The understanding about the difficulties in analyzing emotion made us focus on stages like feature extraction and modeling. During the initial stage, the previous research works were motivated towards emotion dependent parameter analysis by keeping intact the pattern classifiers for emotion recognition [3]. The later works were focused towards development of classifiers and the combination of classifiers Gaussian Mixture Model (GMM) based, Support Vector Machine (SVM) based, Deep Belief Nets (DBN) based [4], combined classifiers [5] by keeping same emotion dependent features [6]. The machine learning and data mining is giving better performance for SER system, but using such methods needs improvement [2].

DBN is one of the deep learning tool used for pattern recognition, voice and speech analysis. The experimental study leads to deeper models and architectures with many visible and hidden layers are disconnected for learning [7]. Many layers and parameters are learnt using deep models [4,8]. The deep learning tools are less used when huge number of parameters are needed for complicated learning process. Training is trapped at local minima and it is time consuming when layers are increased [9]. Achieving acceptable results is difficult. The tool to deal with such problem is DBN by creating deeper networks using many hidden layers [4]. DBN can be used in learning feature and classification. Representation of data is significant in machine learning. So, work done for processing features, extracting features and learning features must be taken more into concern [10]. The feature learning and machine learning can be done using available emotion databases.

Speech emotion database consist of two types namely: simulated emotion speech database and spontaneous emotion speech database. In terms of simulated emotion speech database, the expression conveyed by a subject is generation of various singular feelings of the individuals. However, the speaker has prior knowledge of emotion on which speech elicitation is recorded. Spontaneous emotion contains multiple emotion recorded from a real situation or a genuine conversation. The simulated emotion database is considered due to limited spontaneous emotion database and also not cost effective to develop. The available simulated emotion databases are German emotional speech database(EmodB) and speech under simulated and actual stress (SUSAS) database are the databases available popularly in simulated emotion database. SUSAS database contains the sample emotions of the excluded words elicited by different speakers. SUSAS database has lack of speech recording of continuous sentences. The database utilized here is Classical German emotion database(EmodB) for continuous sentences. The SER system described by Zhou et al., utilizes German Berlin Speech Emotion database by automated feature extraction. The performance acquired using DBN was 65%. The German(EmoDb) is a popular and publicly available emotion database. The motivation to use this database is due to previous experimental observation made in terms of classifiers and accuracy [11]. Further more, the empirical observation described by Pravena et al. [12] on emotionally biased utterance shows higher performance compared to the emotionally neutral utterance for recently developed simulated emotion for Tamil language. The German(EmodB) database and Tamil database are applied in GMM-DBN system, and empirically studied. Speaker emotion recognition is classified into: (1) Speaker Independent and (2) Speaker Dependent. Speaker Independent system does not have a pre-training system for SER. In Speaker dependent, the SER is developed with a training of speakers voice beforehand. The SER used here is for Speaker dependent systems.

The study of the performance for SER using the proposed model are comparable to state of the art classifiers.

This paper discusses on the Modeling of SER system using the available speech emotion database in Sect. 2. The experimental results and its performances are discussed in Sect. 3. The conclusion and future work are discussed in Sect. 4.

2 Emotion Database

2.1 German (EmodB) Speech Emotion Database

EmodB is one of the publicly available popular classical emotion database. The data type available are Speech wave and Electroglottograph(EGG). The speech waves are considered and the total emotion utterances are 535 recorded from five female and five male speakers. Each subject speaks about ten different sentences in seven different emotions anger, happy, sad, neutral, boredom, disgust and fear, out of which only the first 5 emotions are considered for developing the model, sampling frequency is 16 Khz.

2.2 Tamil Speech Emotion Database

The established database consist of three language (Tamil, English and Malayalam) and contains emotions namely anger, happy, sad and neutral utterances. The Tamil database consist of 11 speakers out of which six were female and five were male. The database contains 220 utterances, present for each emotion (anger, happy, sad, neutral) of 9680 sentences(4 emotions * 11 speakers * 220 sentences). The signal recording is on dual channel with a sampling frequency of 48 Khz down sampled at 16 KhZ. The speech and EGG was recorded simultaneously. The present work deals with the speech wave.

3 Modeling of Speech Emotion Recognition System

3.1 MFCC-GMM System for Emotion Recognition

Weighted sum of Gaussian component densities are the representation of parametric probability density function. GMMs are commonly used as a features in a bio-metric system, such as vocal-tract related spectral features in a SER system [13]. Initially the work is to develop a MFCC-GMM system for the existing

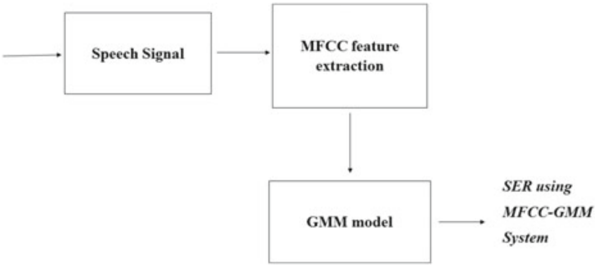


Fig. 1. MFCC-GMM system

German (EmodB) and Tamil speech emotion database. The Gaussian mixture model is a weighted sum of N component Gaussian densities as given by the equation,

$$p(y/\lambda) = \sum_{i=1}^{N} w_i \; g(y/\mu_i, \; \sum{}_i) \tag{1}$$

where y is the features(D-dimensional continuous valued data vector), the mixture weights is $w_i, i = 1, \ldots, N$ and the component Gaussian densities

$$g\left(y/\mu_{i,}\Sigma_i\right), i = 1, \ldots.., M \tag{2}$$

D-variate Gaussian function for each component density is,

$$g\left(y/\mu_{i,}\Sigma_i\right) = \frac{1}{(2\pi)^{D/2}|\Sigma_i|^{1/2}} \exp\left\{-\frac{1}{2}(y-\mu_i)'\Sigma_i^{-1}\left(y-\mu_i\right)\right\} \tag{3}$$

The proposed model here consist of state of the art Mel frequency cepstral coefficients(MFCCs). The MFCC 39 dynamic coefficients are extracted frame by frame from each speech signal. The frame is divided in terms of 20 ms frame with an overlap of 10 ms from each speech utterance, using Hamming window. Spectral leakage is avoided using Hamming window [12]. The GMM model is developed using 256 GMM for German(EmodB) database and 1024 GMM for Tamil database using Expectation maximization algorithm for all the emotions by taking these 39 dynamic MFCC coefficients consist of 13 MFCC features, 13 velocity(Δ) and 13 acceleration($\Delta - \Delta$) coefficients (Fig. 1).

3.2 GMM-DBN System for Emotion Recognition

The minimum distance between the distribution of features for each utterance with respect to each emotion model is averaged with over all emotion is represented as BoF using histogram plot. The BoF is passed in to Deep Belief Network (DBN) as a training vector for emotion recognition (Fig. 2).

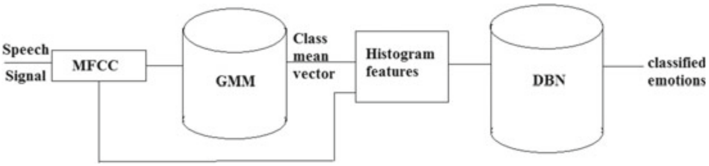


Fig. 2. GMM-DBN system

Experimental Analysis for GMM-DBN System

Let us consider the speech signals (S1, S2, ... Sn), n represents the total number of speech signals. The matrix contains the 39 dynamic MFCC coefficient vector frames F_i (f_1^i............f_{39}^i), i represents the total number of frames for each speech emotion file.

$$S_1 = \begin{bmatrix} f_1^1 \ldots\ldots\ldots\ldots f_{39}^1 \\ f_1^2 \ldots\ldots\ldots\ldots f_{39}^2 \\ . \qquad\qquad . \\ . \qquad\qquad . \\ f_1^n \ldots\ldots\ldots\ldots f_{39}^n \end{bmatrix} \quad \ldots\ldots \quad \ldots\ldots\ldots \quad \ldots\ldots\ldots \quad S_n = \begin{bmatrix} f_1^1 \ldots\ldots\ldots\ldots f_{39}^1 \\ f_1^2 \ldots\ldots\ldots\ldots f_{39}^2 \\ . \qquad\qquad . \\ . \qquad\qquad . \\ f_1^n \ldots\ldots\ldots\ldots f_{39}^n \end{bmatrix}$$

The minimum distance is calculated for each frame of the speech emotion signal with respect to N-Gaussian mixture emotion (c1, c2, c3, c4) and averaged with over all N-Gaussian mixture emotion is represented using histogram (Fig. 3).

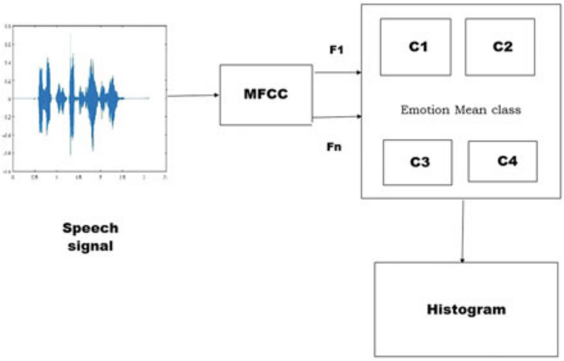


Fig. 3. Histogram

Deep Belief Network

DBN is a deep neural network consist of many hidden layers. DBN allows each RBM model in the sequence to receive different representation of the data. The proposed method is used to develop training in a supervised way [4]. The Restricted Boltzman machine network consist of set of visible units $v \in \{0, 1\}^{n_v}$

and a set of hidden units $h \in \{0,1\}^{n_h}$ where n_v and n_h are the number of visible units and hidden units respectively. The connection between visible and hidden layers are disconnected as shown in Fig. 6.

The different types of RBMs can be modeled. The types defined are generative(data without labels) and discriminative(data with class lables). The sampling methods available are Gibbs, PCD(Persistent Contrastive Divergence), CD(Contrastive Divergence) and FEPCD(Free Energy in Persistent Contrastive Divergence) [14]. The hidden units are independent units due to the disconnection between them, the computation is by giving the training data v, the binary state h_i is set to 1 for each unit i and its probability is given by

$$P\left(h_i = 1/v\right) = g\left(b_i + \sum_j v_j w_{ji}\right) \tag{4}$$

where g(x) is log sigmoid function expressed as $g\left(x\right) = 1/\left(1 + \exp\left(-x\right)\right)$. The connection is not present in between hidden and visible units also, the computation is unbiased. The visible units sample state given the hidden unit vector is computed by,

$$P\left(v_j = 1/h\right) = g\left(a_j + \sum_i h_i w_{ij}\right) \tag{5}$$

Computation is difficult with large running time. So, the CD method is used [15]. In this method, the initialization of visible unit is made with respect to training data. The binary hidden units computation is according to the Eq. (4), the determination of binary hidden unit states leads to computation of v_j values according to the Eq. (5). There are some disadvantages noted that it is not exact due to imperfect gradient computation. This problem is overcome by using PCD method [4]. In this method, it uses the last updated step from the last chain state whereas in the CD method it uses training data as initial value for visible units. The imperfect gradient computation is reduced by using PCD. The FEPCD needs to run many times to obtain appropriate samples from the model and it is impossible. The PCD is being applied here in construction of the RBM layers.

4 Results and Discussion

The German speech emotion recognition for the proposed GMM-DBN system gives the characteristic of performance as shown in Fig. 4. The default 3 RBM based DBN model for which the performance is about 77.27% for PCD binary. The RBM layers are increased and through empirical observation the DBN architecture is constructed as shown in Fig. 5, and performance is about 78.45% as shown in Table 2 The proposed GMM-DBN system performance rate is increased by 5% with comparison to the conventional GMM model is tabulated in Table 1.

Similarly the recent developed simulated emotion for Tamil language is applied in the proposed GMM-DBN system gives the characteristics of performance as shown in Fig. 6. The empirical observation shows that the sampling method Persistent Contrastive Divergence for the unit binary gives maximum performance with comparison to other parameters(CD, FEPCD), for both the databases by applying in the proposed GMM-DBN system. The default 3 RBM based DBN model performance is about 80.74%, comparable to state of the art classifiers.

Table 1. Comparison of MFCC-GMM system and GMM-DBN system performance for EmodB speech emotion database

Type	Performance
MFCC-GMM	73.28%
GMM-DBN	78.45%

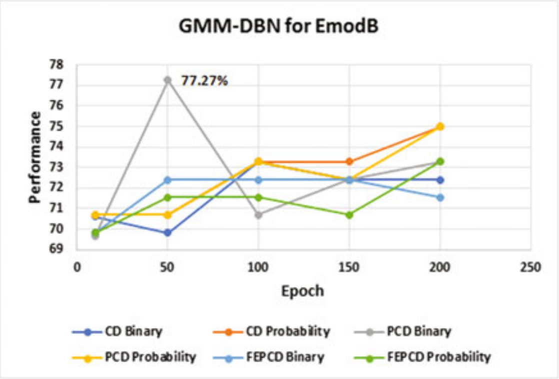


Fig. 4. Performance graph for GMM-DBN system

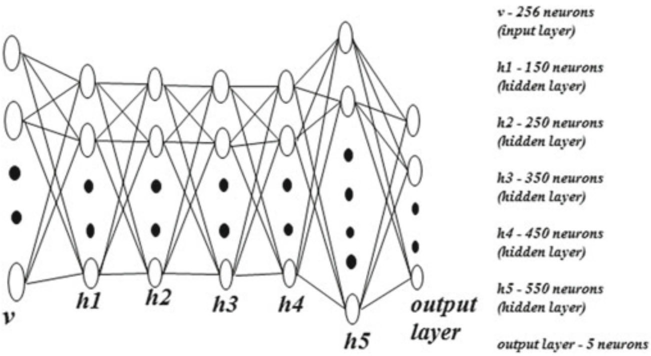

v
h1
h2
h3
h4
h5
output layer
v - 256 neurons (input layer)
h1 - 150 neurons (hidden layer)
h2 - 250 neurons (hidden layer)
h3 - 350 neurons (hidden layer)
h4 - 450 neurons (hidden layer)
h5 - 550 neurons (hidden layer)
output layer - 5 neurons

Fig. 5. DBN architecture for German(EmodB) database

Table 2. GMM-DBN system performance for EmodB database using different RBM layers

Layers	Performance
3	77.27%
5	78.45%

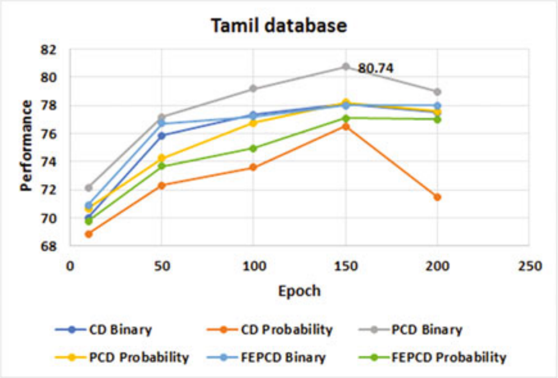


Fig. 6. Performance graph using Tamil speech emotion database for GMM-DBN system

5 Conclusion and Future Work

The proposed work here studies about the effectiveness of GMM-DBN system by testing on the Classical German database(EmodB). The empirical observation shows an emotion recognition performance rate of 5% increase than the conventional GMM model. The GMM-DBN system is also tested with the recently developed simulated emotion database for Tamil language and shows a comparable result for emotion recognition. The effectiveness of the proposed model is clearly observed in EmodB. The characteristics of the GMM-DBN system is empirically studied by various parameters and increase in RBM layers. The present study is used on Speaker dependent. However, the future work needs to concentrate on Speaker independent. Additionally, the present work needs to be improved for other Indian languages.

References

1. Schuller, B., Reiter, S., Muller, R., Al-Hames, M., Lang, M., Rigoll, G.: Speaker independent speech emotion recognition by ensemble classification pp. 864–867 (2005)
2. Scherer, K.R.: Vocal affect expression: a review and a model for future research. Psychol. Bull. **99**(2), 143 (1986)

3. Govind, D., Joy, T.T.: Improving the flexibility of dynamic prosody modification using instants of significant excitation. Circ. Syst. Sig. Process. **35**(7), 2518–2543 (2016)
4. Keyvanrad, M.A., Homayounpour, M.M.: A brief survey on deep belief networks and introducing a new object oriented toolbox (deebnet). arXiv preprint arXiv:1408.3264 (2014)
5. Ververidis, D., Kotropoulos, C.: A state of the art review on emotional speech databases. In: Proceedings of 1st Richmedia Conference, pp. 109–119. Citeseer (2003)
6. Williams, C.E., Stevens, K.N.: Emotions and speech: some acoustical correlates, vol. 52, pp. 1238–1250. ASA (1972)
7. Erickson, D.: Expressive speech: production, perception and application to speech synthesis. Acoust. Sci. Technol. **26**(4), 317–325 (2005)
8. Salakhutdinov, R., Hinton, G.: Deep Boltzmann machines. In: Artificial Intelligence and Statistics, pp. 448–455 (2009)
9. Liu, Y., Zhou, S., Chen, Q.: Discriminative deep belief networks for visual data classification. Pattern Recogn. **44**(10), 2287–2296 (2011)
10. Rong, J., Li, G., Chen, Y.-P.P.: Acoustic feature selection for automatic emotion recognition from speech. Inf. Process. Manage. **45**(3), 315–328 (2009)
11. Altun, H., Polat, G.: On the comparison of classifiers performance in emotion classification: critiques and suggestions. In: 2008 IEEE 16th Signal Processing, Communication and Applications Conference, SIU 2008, pp. 1–4. IEEE (2008)
12. Pravena, D., Govind, D.: Development of simulated emotion speech database for excitation source analysis. Int. J. Speech Technol. **20**, 327–338 (2017)
13. Reynolds, D.: Gaussian mixture models. In: Encyclopedia of Biometrics, pp. 827–832 (2015)
14. Tieleman, T.: Training restricted Boltzmann machines using approximations to the likelihood gradient. In: Proceedings of the 25th International Conference on Machine Learning, pp. 1064–1071. ACM (2008)
15. Carreira-Perpinan, M.A., Hinton, G.E.: On contrastive divergence learning. In: AISTATS, vol. 10, pp. 33–40. Citeseer (2005)

Unsupervised Auditory Saliency Enabled Binaural Scene Analyzer for Speaker Localization and Recognition

R. Venkatesan and A. Balaji Ganesh(✉)

Electronic System Design Laboratory, TIFAC-CORE,,
Velammal Engineering College, Chennai, India
venky88an@gmail.com, abganesh@velammal.edu.in

Abstract. The paper presents an unsupervised binaural scene analyzer that is capable of performing simultaneous operations, such as localization, detection and recognition of target speaker amidst in a reverberant and noise inferential environment. The proposed technique includes three main stages: sound source localization, speaker recognition and auditory saliency-based indexing system. The stage one involves the localization of target speaker by incorporating the binaural cues into Azimuth dependant GMM-EM classifier. During the second stage, the study proposes a Gabor-Hilbert Envelope Coefficient (GHEC) based spectro-temporal extractor as an efficient speaker recognition method which shows better robustness with minimum computational complexity. The Hilbert Envelope provides relevant acoustic information and also it improvises the performance of speaker identification process in different reverberant environments and SNR values. Later in the third stage, the auditory saliency based diarization is proposed as a process of indexing the speech contents based on the image identity of the target speaker. The proposed system may be used to catalogue the entire speech content with the corresponding image of the target speaker that finds wide range of applications, including teleconference, hands-free communication and meeting hall content localization and fast audio retrieval in a repository.

Keywords: Binaural cues · Computational auditory scene analysis · Automatic speaker recognition · GHEC · Shortterm fourier transform · Auditory saliency

1 Introduction

Human auditory scene analysis has the ability to localize, recognize as well as to segregate target sound source from complex acoustic mixtures [1]. Almost, all the auditory computational algorithms are predominantly inspired by the features of human hearing system which precociously handles several functions, such as speech recognition, pitch and distance perception, sound localization, noise suppression and assessment of qualitative characteristics of sound signals [1, 2]

In spite of the advancements in computing technologies the adaptation of human auditory model into computer algorithms especially when multiple target speakers are gathered in a noisy environment is still considered as an extremely challenging task [2].

S.M. Thampi et al. (eds.), *Advances in Signal Processing and Intelligent Recognition Systems*, Advances in Intelligent Systems and Computing 678,
DOI 10.1007/978-3-319-67934-1_30

The sound source localization promotes many applications, including hearing prostheses, reproduction of spatial sound and autonomous robots. It is done by estimating interaural time difference (ITD) and interaural level difference (ILD) between two ears [1, 2].

Speaker diarization is a process through which speech input streams are partitioned with appropriate speaker identity information. The applications of diarization, includes speech and speaker indexing, document content structuring, speaker recognition, speaker attributed speech to text transcription, speaker recognition in the presence of multiple or competing speakers [3]. Recently, significant research works on speaker recognition are done using the algorithms, such as GMM-UBM, Joint Factor Analysis and i-vector based techniques [2, 4, 5]. In general, Hilbert Envelope [6] is performed by creating the analytic signal of the input streams by using Hilbert transform. The Hilbert transform helps to analyse the instantaneous frequency content and power of the signal. This Envelope extraction does not significantly affect the amplitude of the signal [6].

The objectives of this study are to localize, detect and recognize the speaker amidst reverberation and noise interferences by analyzing the input binaural speech signal. Also, a visual-auditory saliency based speaker diarization system is demonstrated as fast content browsing with less computational complexity. The target source localization is done by applying GMM-EM classifier on binaural cues which are obtained from Gammatone filter bank. The study proposes a spectro-temporal pattern extractor which is referred as GHEC and it shows good robustness against various noises and reverberant conditions. For the speaker detection and recognition both i-vector extraction techniques along with channel compensation technique and also GMM-UBM are applied. In this work, the usefulness of the auditory saliency map is incorporated by extracting the auditory features, such as pitch, envelope, intensity, orientation and color and combining the conspicuity maps into a final saliency map as bottom-up approach. The DT-CWT is employed to extract salient discriminant information of each speakers and also it helps to obtain invariant features of acoustic mixture. From the results, it is found that auditory saliency process has been fast enough to select the predominant event in an acoustical scene than spectrogram based browsing technique.

2 Model Architecture

The block diagram representation for the experiments carried out in this study which is shown in Fig. 1.

2.1 Speaker Sound Source Localization

2.1.1 Binaural Localization Analysis

The present study is inspired by the human cochlea which is known for its frequency selectivity. In the literature of human hearing, the cochlea and inner hair cell in the auditory system are modelled by using digital Gammatone filters [2, 8]. The speech signals are decomposed into auditory channels which acts as a bank that consists of (N = 32) fourth-order and phase-compensated filters. In addition, each Gammatone filter bank channel is scaled with a specific gain to model the frequency response of the

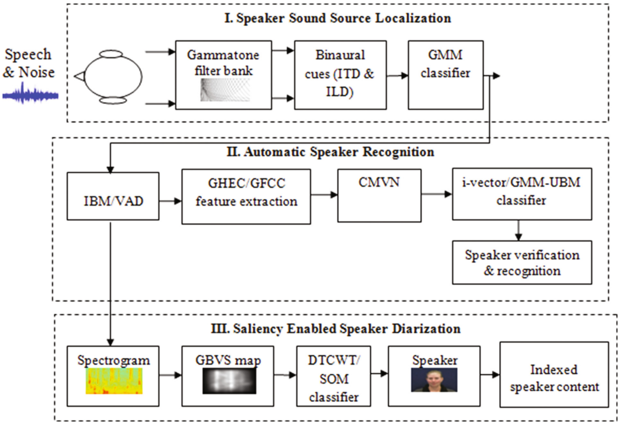

I. Speaker Sound Source Localization
Speech & Noise
Gammatone filter bank
Binaural cues (ITD & ILD)
GMM classifier
II. Automatic Speaker Recognition
IBM/VAD
GHEC/GFCC feature extraction
CMVN
i-vector/GMM-UBM classifier
Speaker verification & recognition
III. Saliency Enabled Speaker Diarization
Spectrogram
GBVS map
DTCWT/ SOM classifier
Speaker
Indexed speaker content

Fig. 1. Block diagram representation of unsupervised speaker localization, recognition and transcription system.

middle ear [2, 8]. The Eq. (1) represents the Gammatone filter in time domain as an impulse response function.

$$g(t) = at^{\mathrm{n}-1}\cos(2\pi \mathrm{ft} + \emptyset)e^{-2\pi \mathrm{bt}} \tag{1}$$

The parameters of the Gammatone filter are, n is the order, a amplitude, b (in Hz) bandwidth with respect to the equivalent rectangular bandwidth (ERB) scale to control the duration of the impulse response function, f (in Hz) denotes the nominal centre frequency of the carrier, ∅ (in radians) the carrier phase that determines the relative position of the fine structure of the carrier to the envelope. The centre frequencies of the filter banks are equally spaced on the ERB scale between 80 Hz and 5 kHz channel-dependent gains [2, 3, 8].

The speech signals arriving at the left and the right ear are decomposed into auditory channels using a fourth-order Gammatone filter bank with 32 auditory channels. The aligned centre frequencies of Gammatone channels are utilized to compensate the group delays. The auditory signals are validated as rectangular window of 20 ms at a sampling frequency of 44.1 kHz corresponding to 882 samples with an overlap of 50% between the successive frames.

The binaural feature extraction consists of estimation of both ITD and ILD by using cross-correlation analysis in time domain and by calculating the energy per frame, respectively. The 3-D Azimuth- dependent likelihood function is created immediately after the estimation of ITD and ILD cues.

In general, ITD cues provide a good impact for the auditory signals with low level frequencies. ITD searches for the main peak in the Generalized Cross Correlation (GCC) between the left and the right ear. The integration of ILD-ITD is done by incorporating azimuth-dependent Gaussian Mixture Model (GMM). The ITD-ILD spatial log-likelihood function is generated for azimuth angles from -180^0 to $+180^0$ and spaced by 5^0 in between. GMM is trained for the target signals at both higher and lower frequencies [2]. The log likelihood computation is extended for all the 32 Gammatone channels. The mixture density used to determine the maximum likelihood function is defined as,

$$p(x/\lambda) = \sum_{i=1}^{M} w_i p_i(x) \tag{2}$$

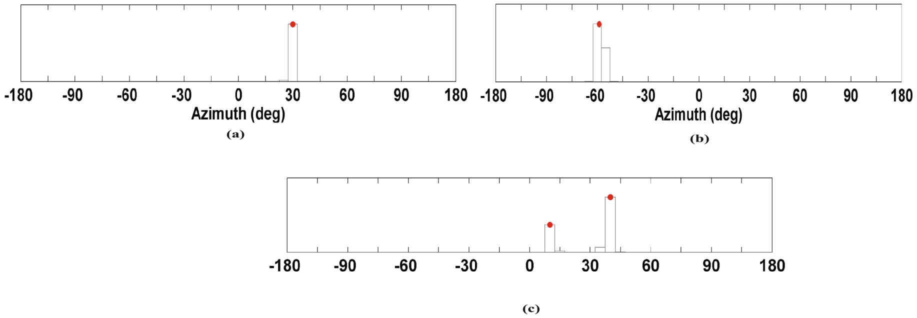
Where,

$$p_i(x) = \frac{1}{(2\pi)^{\frac{D}{2}} \left|\sum_i\right|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(x-\mu_i)' \sum\nolimits_i^{-1} (x-\mu_i)\right\}$$

The Uni-modal Gaussian densities depend on the mean μ_i, $D \times 1$ vector, and a $D \times D$ covariance matrix, $\sum_i$ and w_i denote the mixture weights. The parameters of the density model are defined as $\lambda = \left\{w_i, \mu_i, \sum_i\right\}$. Log-likelihood is denoted as,

$$\log p(X/\lambda) = \sum_{i=1}^{T} \log p\,(x_i/\lambda) \tag{3}$$

Where, $p(x_i/\lambda)$ is manipulated using the Eq. (2) (Fig. 2).



(a)

(b)

(c)

Fig. 2. (i) shows the single target speaker localization at 30^0 (ii) shows the single target speaker localization at 60^0 (iii) shows the two target speakers localization at 10^0 and 40^0 consequently.

2.2 Automatic Speaker Recognition

2.2.1 Monaural IBM Estimation

The level of complexity in recognizing the speaker depends upon two important factors, namely the number of target speakers and the noise sources in the reverberant environment. The system uses the combination of GHEC and GFCC monaural features for speech segregation.

2.2.2 Gabor Hilbert Envelope Coefficients

The current system proposes Gabor Hilbert Envelope Coefficient (GHEC) monaural feature in which Gabor filters are convolved with Hilbert Envelope in order to obtain the improved performance of conventional speaker recognition system. The GHEC based feature extraction process is illustrated in the Fig. 3. The modules that are involved in GHEC The information about spectral, temporal and spectro-temporal components are extracted through Gabor features by using set of Gabor filters. Further, the Gabor features [9] are utilized to improvise the classification rate. The extracted feature components are dependent of the output of Gabor filters and eventually its convolution with Hilbert Envelope. The output of Hilbert transform, $H_{tr}(s,i)$ contains both real and Hilbert transformed part and it is used for the computational process of envelope, $H_e(s,i)$.

$$H_{tr}(s,i) = G_{cr}(s,i) + iG'_{ci}(s,i) \tag{4}$$

Where, $G_{cr}(s,i), G'_{ci}(s,i)$ are the real and Hilbert transformed real signal, respectively and i, is an imaginary unit. The Hilbert Envelope $H_{en}(s,i)$ is obtained by using the Eq. (5).

$$H_{en}(s,i) = G_{cr}(s,i)^2 + G'_{ci}(s,i)^2 \tag{5}$$

The Hilbert Envelope is smoothed through exploiting low pass filter with cut-off frequency of 20 Hz in order to remove the redundant unwanted higher frequency components. The smoothed Envelope, $H_s(s,i)$ is grouped into 25 ms duration with a skip rate of 10 ms. Further, the discontinuities at the edges of each frame are diminished by the process of Hamming window. The sample means are estimated as

$$N(t,i) = \frac{1}{N}\sum\nolimits_{s=0}^{N-1} w(s)\ H_s(s,i) \tag{6}$$

Where, w(s) is a Hamming window. The natural logarithm is applied on the estimated resultant parameter, N(t,i) which is used here as a channel normalization technique in order to bring human perception of loudness as well as to compress the dynamic range. As a final step, Discrete Cosine Transform (DCT) is used to perform two functions, namely conversion of spectral features into Cepstrum and also to decorrelate various overlapped feature dimensions. The first and second Cepstral derivatives are calculated and added to the features in order to capture various 57 dimensional dynamic patterns.

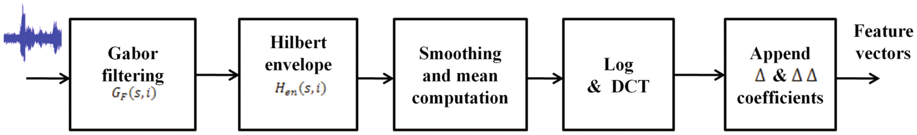


Fig. 3. shows the extraction process of monaural GHEC feature from the input speech signal.

2.2.3 Speaker Identity Recognition

2.2.3.1 i-Vector Based Speaker Recognition System

The experiment exploits 57-dimensional Gabor Hilbert envelope features with the appended delta coefficients are extracted as discriminant acoustic features from speech material for i-vector based speaker verification techniques. The channel/session variability is referred as mismatch between trained and test data and it is promoted by numbers of factors, including noise sources, variations in voice of the speaker and environmental conditions. The channel and session variability can be compensated through significant techniques such as With-in Class Covariance Normalization (WCCN), Linear Discriminative Analysis (LDA) and Source-Normalized Weighted Linear Discriminant Analysis (SN-WLDA) [10]. The Joint Factor Analysis (JFA) comprises both speaker or Eigen-voices and channel or Eigen-channel variability in two different spaces. The speaker-dependent GMM super vector, k consists of separate speaker and channel dependent components, S_d and C_d respectively are given as,

$$k = S_d + C_d \tag{7}$$

Where $S_d = m + Vy + Dz$; $C_d = Ux$; where, m represents session and speaker independent super vector extracted from UBM (Universal Background Model); x, y and z denoted as tspeaker and session dependent factors in their respective subspace. V and D specify the speaker subspace whereas U represents session subspace. In the recent literature, i-vector extraction along with Probabilistic Linear Discriminant Analysis (PLDA) has been experimentally proved as an improvised and efficient technique than conventional JFA and SVM for the speaker verification [10]. Also, the GMM super vectors are represented by a total-variability space unlike the separate subspaces for speaker and channel variability in JFA [4, 5]. The feature warping is considered as a normalization technique and it is carried out to enhance the robustness of the system as well as to minimize the mismatches during classification.

In this study, the speech samples under environmental conditions such as noisy or reverberant are trained by using pooled total-variability space approach. The resultant output from total-variability space is utilized by GPLDA (Gaussian Probabilistic Linear Discriminant Analysis) classifier. The total-variability space for speaker and channel dependent GMM super vector, k is defined as,

$$k = m_s + T_r w \tag{8}$$

Where, m_s is the session and speaker independent UBM super-vector, T_r is a low-rank matrix representing the primary directions of variability across all development data and w denotes, the independent normal distributed random vector with parameter N (0,1).

WCCN along with LDA: WCCN (Within Class Covariance Normalization) is explored to compensate dimensions of high within-class variance [10]. The major demerit of WCCN is it additionally removes the dimensions of between-class variance while it reducing the dimensions of high with-in class variability. The issue can be overcome by combining WCCN along with LDA (Linear Discriminant Analysis) for the transformation of total-variability.

As described, LDA is responsible to produce reduced set of axes A through Eigen-value decomposition where as WCCN transformation matrix (B_t) is derived by following Cholesky decomposition of $B_t B_t^T = W^{-1}$ where W is computed by using,

$$W = \frac{1}{S}\sum_{s=1}^{S}\sum_{i=1}^{n_s}\left(A_s^T\left(w_i^s - \overline{w}_s\right)\right)\left(A_s^T\left(w_i^s - \overline{w}_s\right)\right)^T \tag{9}$$

The final WCCN [LDA] is obtained by computing,

$$\mathrm{WCCN[LDA]} = B_t^T A_s^T \mathrm{W} \tag{10}$$

GPLDA: In the earlier literature, significant work has demonstrated the PLDA based i-vector speaker recognition technique by creating session and speaker variability within the i-vector space, effectively [10]. Recently, length normalised GPLDA approach is introduced by converting the i-vector feature behaviour from heavy-tailed to Gaussian. The GPLDA based i-vector speaker recognition technique involves extraction of i-vector, session variability compensation, likelihood ratio scoring and the results of this technique along with other compensation methods based on GHEC feature are shown in Table 1. A speaker and channel dependent length-normalised i-vector, w can be defined as,

$$w = \overline{w} + E_1 x_1 + E_2 x_2 + \gamma_r \tag{11}$$

Where, γ_r is residual with mean zero, E_1 and E_2 are the Eigen voice matrix and Eigen channel matrix, respectively.

GPLDA scoring is computed by utilising batch likelihood ratio and it provides ratio between two i-vectors of target and test speakers. It is defined as,

$$\ln = \frac{P\left(w_{target,} w_{test}|H_1\right)}{P\left(w_{target}|H_0\right)P\left(w_{test}|H_0\right)} \tag{12}$$

Where, H_1: The speakers are same, H_0: The speakers are different.

Table 1. Comparative analysis of channel compensation techniques for different truncation of testing and training utterance under reverberant condition (RT = 0.38 s).

Utterance size (training-testing)	JFA system		WCCN-LDA		GPLDA	
	EER%	DCF	EER%	DCF	EER%	DCF
Full (5 s)-2 s	13.02	0.049	11.76	0.043	13.16	0.045
Full (5 s)-4 s	9.17	0.045	8.53	0.036	9.33	0.032
Full-Full	8.07	0.032	7.97	0.030	8.03	0.033

2.2.3.2 GMM-UBM Model Based Speaker Verification and Recognition

The Automatic speaker recognition is a process which helps to identify an individual person by analyzing the spectral content of his/her speech sources. In general, the performance of this process is affected as when reverberation time increases. In a reverberant room environment, the source signal reaches the target after experiencing series of reflections and diffractions. A typical speaker recognition system involves two phases, namely training and classification. The training phase is involved with delta coefficients and eventually the trained data during classification phase is processed by using GMM-UBM [2] with 57 dimensional GHEC features.

In the literature, the Mel-frequency Cepstral Coefficients (MFCCs) as well as Gabor Filter-Bank Features (GBFB) have been used extensively in many speech processing applications, including speech, emotion and language recognition as well as for the speaker recognition [9]. Here, Gabor filters are convolved with Hilbert Envelope and proposed as GHEC for the speaker recognition. The performance of GHEC is compared with various known algorithms, including MFCC and GFCC. The Cepstral Mean and Variance Normalization (CMVN) [6] are applied on the Cepstral coefficients obtained from GHEC in order to further improvise the robustness of automatic speaker recognition process. The GHEC monaural is computed from extracted cepstrum by averaging the acoustic signals of left and right ear, whereas GHEC binaural is done by selecting better ear based on the feature vector with higher SNR values. In this study, a 64 mixture components of UBM is trained during GMM-UBM based speaker recognition phase with EM algorithm by using speech data that are collected from large numbers of speakers (both male and female) with the relevance factor of 19.

In the recognition mode, the Log Likelihood Ratio (LLR) score is estimated for the feature vectors, $X = \{\lambda_{\text{impost},}\lambda_{\text{tar}}\}$ and LLR is derived as,

$$\text{LLR}_{\text{score}} = \log(X|\lambda_{\text{tar}}) - \log(X|\lambda_{\text{impost}}) \tag{13}$$

Leas Where, γ_{tar} are the utterances which are related to target speaker and γ_{impost} are the acoustic signal which are related to non-target speaker. The speaker verification is carried out to confirm whether a speech source signal can be accepted or rejected which is based on the predefined threshold θ,

$$\text{LLR}_{\text{score}} = \begin{cases} \geq \theta & \textit{accepted} \\ < \theta & \textit{not}-\textit{accepted} \end{cases} \tag{14}$$

2.3 Saliency Enabled Diarization Process

The proposed framework contains diarization process and it is done by analysing the auditory saliency information of target speaker. The auditory saliency indexing process has various advantages, including fast browsing and efficient content localization. The final auditory saliency map is created by combining the normalised conspicuity map with features, such as intensity, orientation, pitch, envelope and color. Though auditory saliency map consists of both salient sound information and back ground noise, the desired salient point is detected through Inhibition Of Return (IOR) model [11].

2.3.1 Bottom-up Approach Features

On the basis of detecting the salient sound from a scene, it is believed that the general auditory saliency analysis significantly matches with human auditory perception analysis. In this study, the auditory saliency analysis is carried out to extract the various features, including color, intensity, orientation and frequency cues [7, 11]. The developed framework involves a saliency-maximized auditory spectrogram and it is generated by extracting multiple features, such as intensity, color, pitch, envelope and orientation. In general, these features act as a weighted representation of the acoustic signals. Thus, detailed structure of the auditory saliency enabled diarization is shown in Fig. 4. The feature classes, such as color, orientation and intensity compete to give a winning location of salience via centre-surround operations. A complete bottom-up saliency model is based on graph computation and referred as Graph Based Visual Saliency (GBVS) that consists of framework of activation and normalization/combination. It depends on Kullback-Leibler distance between central Gaussian distributions and its surrounding distributions. The spatial scales are obtained by

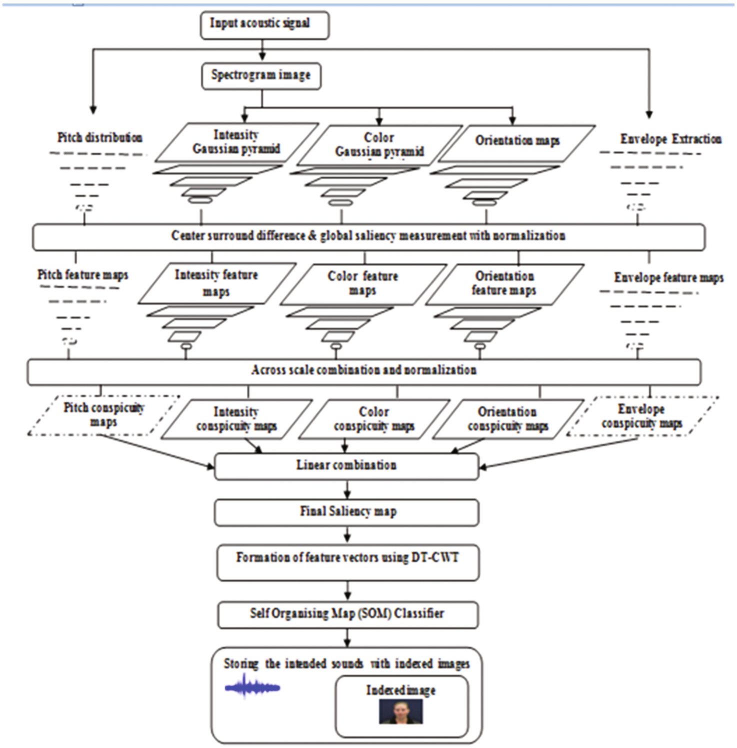

Input acoustic signal
Spectrogram image
Pitch distribution
Intensity Gaussian pyramid
Color Gaussian pyramid
Orientation maps
Envelope Extraction
Center surround difference & global saliency measurement with normalization
Pitch feature maps
Intensity feature maps
Color feature maps
Orientation feature maps
Envelope feature maps
Across scale combination and normalization
Pitch conspicuity maps
Intensity conspicuity maps
Color conspicuity maps
Orientation conspicuity maps
Envelope conspicuity maps
Linear combination
Final Saliency map
Formation of feature vectors using DT-CWT
Self Organising Map (SOM) Classifier
Storing the intended sounds with indexed images
Indexed image

Fig. 4. explains the auditory saliency feature extraction along with unsupervised classification technique and indexing the speakers' content.

progressively sub-sampling the input spectrogram through Gaussian pyramids. The saliency map analysis is involved with two important processes, namely construction of feature map and eventually generation of conspicuity map. The final auditory saliency map and the short time fourier transform of speech source are illustrated in Fig. 5. The construction of feature map involves the processes, including extraction of features for the pyramids and choosing various scales within pyramids for the features. The speaker identification process is further enhanced by incorporating the resultant conspicuity map and it is promoted as fast and efficient browsing system [7].

2.3.2 DT-CWT Based Feature Extraction

In general, the Dual-Tree Complex Wavelet Transform (DT-CWT) is a complex wavelet $((\gamma\!: = \gamma_r(t) + i\gamma_{\mathrm{i}}(\mathrm{t})))$ that generates both real and imaginary parts; where,$\gamma_{\mathrm{r}}(\mathrm{t})$ is approximately analytic and $\gamma_{\mathrm{i}}(\mathrm{t})$ is approximately its Hilbert transform. It has been applied for the creation of de-noising model to extract the invariant features. The DT-CWT is considered as an improvised model than conventional DWT as it provides high degree of shift-invariance and better directionality [12, 13]. The real two dimensional filter banks of DWT contain three high-pass sub-bands with orientations of 0^0,45^0 and 90^0 whereas, two dimensional DT-CWT produces directional sub-bands for each scale oriented at ±15°, ±45° and ±75°. The DT-CWT derives complex coefficients for each directional sub-bands at each scale and their magnitude in which it is given by,

$$\mathrm{s(x,y)} = \sqrt{\mathrm{R}_{\mathrm{rs}}^2 + \mathrm{C}_{\mathrm{rs}}^2} \tag{15}$$

Where,R_{rs} and C_{rs} are the complex coefficients, s refers scale, r refers six sub-bands oriented at ±15°, ±45° and ±75°.

Here, the DT-CWT is used to extract the edge and illumination invariant features from facial/spectrogram image. These two discriminant information are combined in order to improvise the invariant features that results enhanced robustness of the system, especially in varying lighting conditions. The output of DT-CWT [12] is the formation of six sub-band images from which feature vectors are formed by computing statistical measures such as Entropy, percentile kurtosis, percentile skewness, kurtosis, skewness, variance, and standard deviation. The feature vector is used for the Self-Organising Map (SOM) classification phase [7].

2.3.3 Classification Through SOM

The Self-organizing Map (SOM) [7] is a kind of artificial neural network classification algorithms which is used as an unsupervised learning method in order to create a topographically ordered spatial representation of an input pattern. The map maintains topological linkage between the inputs in a way that neighbouring inputs in the input space are mapped to neighbouring neurons in the map space. With the advent of SOM,

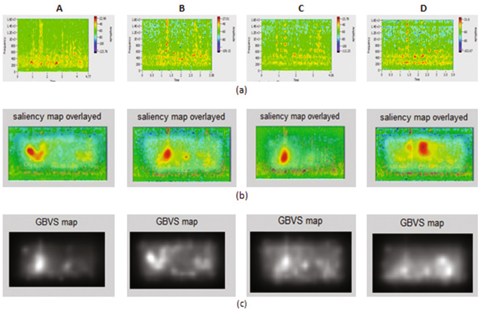


Fig. 5. shows extraction of auditory salient features from auditory spectrogram (a) auditory spectrogram (b) auditory saliency map (c) GBVS map for different target speakers.

the winning neuron is estimated by minimum Euclidean distance between input x_i and weighting vector w_{ij} and minimum Euclidean distance for each cluster j is given by,

$$D(j) = \sum_{i} (x_i - w_{ij})^2 \qquad (16))$$

To update the adjustable weights w_{ij}
$w_{ij}(new) = w_{ij}(old) + \lambda\left(x_i - w_{ij}(old)\right)^2$ Where, λ denotes the learning rate.

3 Results and Discussions

The VidTIMIT [16] dataset consists of video and corresponding audio recordings of 80 volunteers (39 female and 41 male), reciting short sentences. The recording had been done in an office environment using a broadcast quality digital video camera. The NOIZEUS dataset is utilized for the recognition rate estimation of different features paradigm in various noisy conditions. The speech signals are convolved with impulse responses (BRIR) that are obtained from Aachen Impulse Response (AIR) database for different rooms and also from university of surrey [1, 2]. For the speaker recognition experiments, the speech signals from the TIMIT dataset [15] are also utilised.

3.1 GMM-UBM IBM and i-Vector IBM Based Speaker Recognition

The localization of target speaker by using binaural acoustic signals is done by incorporating azimuth-dependent Gaussian Mixture Model (GMM) along with joint ILD-ITD estimation. The performance of GMM-ITD-ILD under various reverberation time periods is compared with both GMM-ILD and GMM-ITD [1, 2]. The performance of proposed GHEC is compared with other standard feature extraction techniques, namely GFCC, RASTA-MFCC, MHEC and GBFB for various SNR values with different noise sources and also for the various reverberant room environments. The GFCC feature extraction [14] is done by using a total number of 64 channel Gammatone filter-banks with central frequencies ranges from 50 Hz to 8000 Hz.

The outputs of rectified filter response are decimated into 100 Hz which yields time frames of 10 milli-seconds.

The MHEC [6] feature extraction is performed by using 24 channels Gammatone filter banks with centre frequencies spaced on Equivalent Rectangular Bandwidth (ERB) scale between 300 and 3400 Hz that are utilised to decompose the speech signal into 24 bands. The Hilbert Envelope is computed along with smoothing and mean computation. Then, first and second derivatives are done and appended to the features in order to design final 36 dimensional MHEC feature patterns. The performance overview of various feature extraction techniques under different non-stationary noisy environments with SNR values are shown in Table 2.

It is understood that the accuracy of speaker identification by using GHEC (57 dimensional features) outperforms RASTA-MFCC (36 dimensional features), significantly for all the noise acoustic signals under various SNR values. Also, it produces comparatively better results than GBFB along with Principal Component Analysis (PCA) that has total length of 39 dimensional features for the various noise signals especially with the SNR value of –5 dBA. The performance results are found to be almost similar for GHEC, MHEC and GFCC for the various noise signals especially with low SNR values. The fluctuations of spectro-temporal pattern of GHEC is affected in some SNR values.

In this study, the text independent GMM-UBM-IBM as well as i-vector-IBM based speaker recognition is validated in both anechoic and reverberant conditions. The feature extractions are done by using both GFCC and GHEC techniques. It should be noted that the GMM-UBM and i-vector based speaker recognition system utilises 64

Table 2. demonstrates evaluation of recognition accuracy of different features paradigm in various noisy conditions.

Babble noise	−5 dBA	0 dBA	5 dBA	15 dBA	30 dBA	Average
GHEC (proposed)	28.05	80.53	86.42	98.36	99	73.65
RASTA-MFCC	25.65	73.34	70.04	96.2	97.3	68.23
GFCC	27.87	70.03	72.44	98	98.16	68.93
MHEC	27.52	75.35	85.52	98.38	99	72.09
GBFB	26.57	74.67	81.63	98	98.32	71.31
Street noise	−5 dBA	0 dBA	5 dBA	15 dBA	30 dBA	Average
GHEC (proposed)	42.24	70.78	84.32	96	98.27	72.68
RASTA-MFCC	35.37	71.32	80.64	94.25	95.16	70.17
GFCC	35.26	65.43	73.23	95	97.06	68.98
MHEC	45.34	71.41	84.57	97.5	97.57	73.46
GBFB	44.27	70.04	82.57	95	96.6	71.97
Car noise	−5 dBA	0 dBA	5 dBA	15 dBA	30 dBA	Average
GHEC (proposed)	51.53	71.33	84.27	93.64	95.43	75.22
RASTA-MFCC	53.34	69.27	72.45	81.32	97.43	69.95
GFCC	52.42	68.15	73.23	91.35	97.16	72.82
MHEC	57.35	70.42	80.15	92.3	94.23	75.54
GBFB	56.67	70.24	80.35	94.57	95.06	75.73

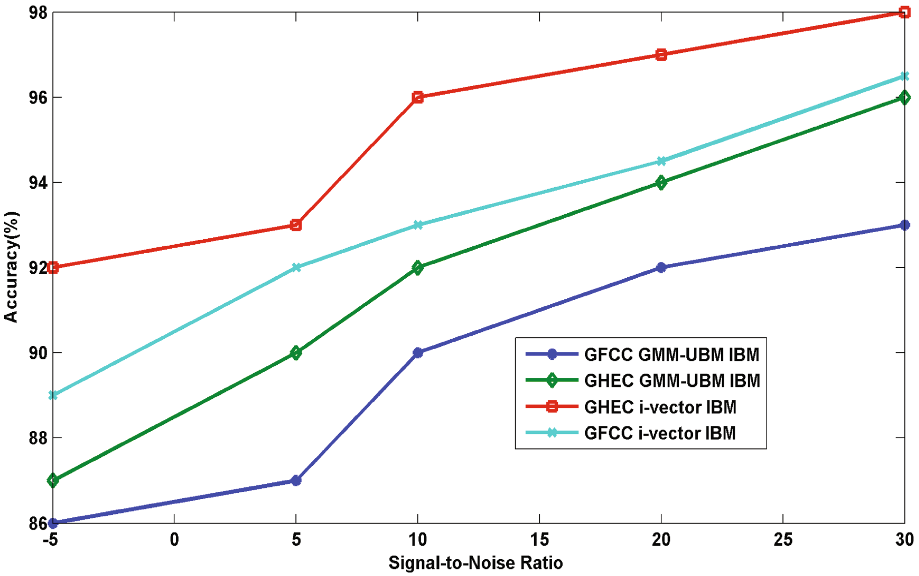


Fig. 6. shows the GHEC and GFCC performance measure under different speaker acoustic models with single noise source (babble)

Gaussian mixture components for training of speech samples. From the results, it is observed that the GHEC provides better performance than GFCC for Babble noise under various SNR values which is shown in Fig. 6.

4 Conclusion

A binaural scene analyzer is demonstrated successfully that proved its ability to localize, identify and recognize the target speaker, simultaneously in both anechoic as well as reverberant environments. The determination of location information of target speaker is considered as an efficient value added technique in conventional speaker recognition system. The work proposes a new feature extraction technique, referred as GHEC and it is implemented by combining Gabor Filter-banks with Hilbert Envelope. The performance of GHEC is validated in different noisy environments with SNR values and shows better results than conventional techniques. The speaker identification (SID) is performed by using numbers of techniques; including the proposed Gabor based Hilbert coefficients along with i-vector/GMM-UBM. The obtained results proved that the extraction of spectro-temporal features from the acoustic mixture can further improvise the robustness. The work also proposes an auditory saliency based fast browsing technique which is demonstrated as an efficient speech content diarization process.

References

1. May, T., van de Par, S., Kohlrausch, A.: A binaural scene analyzer for joint localization and recognition of speakers in the presence of interfering noise sources and reverberation. IEEE Trans. Audio Speech Lang. Process. **20**(7), 2016–2030 (2012)
2. Kohlrausch, A., Braasch, J., Kolossa, D., Blauert, J.: The Technology of Binaural Listening. Springer, Berlin (2013)
3. Anguera, M.X., Bozonnet, S., Evans, N., Fredouille, C.: Speaker diarization: a review of recent research. IEEE Trans. Audio Speech Lang. Process. **10**, 356–370 (2012)
4. Dehak, N., Kenny, P., Dehak, R., Dumouchel, P., Ouellet, P.: Front-end factor analysis for speaker verification. IEEE Trans. Audio Speech Lang. Process. **99**, 1 (2011)
5. Dehak, N., Dehak, R., Glass, J., Reynolds, D., Kenny, P.: Cosine similarity scoring without score normalization techniques. In: Odyssey Speaker and Language Recognition Workshop (2010)
6. Sadjadi, S.O., Hansen, J.H.L.: Mean hilbert envelope coefficients (MHEC) for robust speaker and language identification. Speech Communication **17**, 138–148 (2015)
7. Venkatesan, R., Reeni, J., Balaji Ganesh, A.: A saliency based effective browsing of visual and acoustics. Aust. J. Basic Appl. Sci. **9**(16), 97–103 (2015)
8. Woodruff, J., Wang, D.: Binaural localization of multiple sources in reverberant and noisy environments. IEEE Trans. Audio Speech Lang. Process. **20**(5), 1503–1512 (2012)
9. Schadler, M.R., Meyer, B.T., Kollmeier, B.: Spectro-temporal modulation subspace-spanning filter bank features for robust automatic speech recognition. J. Acoust. Soc. Amer. **131**, 4134–4151 (2012)
10. Kanagasundaram, A., Dean, D., Sridharan, S., Vogt, R.: i-vector based speaker recognition using advanced channel compensation techniques. Comput. Speech Lang. **28**, 121–140 (2014)
11. Hong, T., Kingsbury N., Furman, M.D.: Biologically-inspired object recognition system with features from complex wavelets. In: Proceedings of 18th IEEE International Conference on Image Processing (ICIP), pp. 261–264 (2011)
12. Haifeng, H.: Illumination invariant face recognition based on dual-tree complex wavelet transform. IET Comput. Vision **9**(2), 163–173 (2015)
13. Selesnick, I., Baraniuk, R., Kingsbury, N.: The dual tree complex wavelet transform. IEEE Signal Process. Mag. **22**(6), 123–151 (2005)
14. Zhao, X., Shao, Y., Wang, D.L.: CASA-based robust speaker identification. IEEE Trans. Audio Speech Lang. Process. **20**(5), 1608–1616 (2012)
15. Garofolo, J., et al.: TIMIT Acoustic-Phonetic Continuous Speech Corpus. Linguistic Data Consortium, Philadelphia (1993)
16. Sanderson, C., Lovell, B.C.: Multi-region probabilistic histograms for robust and scalable identity inference. Lecture notes in computer Science (LNCS), Vol. 5558, pp. 199–208 (2009)

Applying Machine Learning Techniques for Sentiment Analysis in the Case Study of Indian Politics

Annapurna P. Patil(✉), Dimple Doshi, Darshan Dalsaniya, and B.S. Rashmi

Department of Computer Science and Engineering,
Ramaiah Institute of Technology, Bengaluru 560054, India
annapurnap2@msrit.edu, djdimplejd@gmail.com,
darshandalsaniya@gmail.com, rashshobli@gmail.com

Abstract. In the recent era, humans have become detached from their surroundings, immediate peers and more addicted to their social media platforms and micro-blogging sites. Technology is digitalizing at a very fast pace and this has led to man being social, but only on technological forefront. Social media platforms like twitter, facebook, whatsapp, instagram are in trend. In our paper, we will concentrate on data generated through Twitter (tweets). People express their opinions, perspectives within a 140 character tweet, which is subjective. We try to analyze their emotion by tweet classification followed by sentiment analysis. On an average, with 328 million Twitter users, 6000 tweets are generated every second. This tremendous amount of data can be used to assess general public's views in economy, politics, environment, product reviews, feedbacks etc. and so many other sectors. Here, we take into account the political data from Tweets. The data obtained can be images, videos, links, emoticons, text, etc. The results obtained could help the government function better, improve their flaws, plan out better strategies to empower the nation.

Keywords: Subjective data · Polarity · Machine learning · Twitter · Sentiment analysis

1 Introduction

With immense progress in technological development, social networking platforms are very popular. The tweets that are collected are classified by its positivity, neutrality, negativity. In recent times, Barack Obama's historic win was credited to similar online strategies during US polls. Donald Trump's victory was also on similar lines. Users tweet using '#'- hashtag at the beginning of the statement. #MakeAmericaGreatAgain or #NaMoForPM were trending hashtags for extending support to their desirable candidates. This shows that an accurate, well-defined data analysis can be used to predict Presidential Elections with high reliability. Sentiment analysis is one such approach. It obviously cannot guarantee the results, but it just predicts the most probable results based on the attitude of the tweets. Thus, sentiment analysis can be defined as analyzing the emotions, opinions from unstructured data with help of

S.M. Thampi et al. (eds.), *Advances in Signal Processing and Intelligent Recognition Systems*, Advances in Intelligent Systems and Computing 678,
DOI 10.1007/978-3-319-67934-1_31

computational techniques like text mining, natural language processing, text classification. The end results would determine the polarity of the entity, where an entity is an individual, topic or event. Also, our paper will show a comparison of the various classifiers based on their accuracies. With evolving algorithms and better classification techniques the accuracy of prediction will shoot up in future.

This paper is organized as: Sect. 2 is about Literature survey already done in this area, Sect. 3 deals with the overall methodology, Sect. 4 deals with the various approaches and classification methods to perform sentiment analysis, Sect. 5 deals with Results and accuracy, Sect. 6 deals with conclusion and future scope.

2 Literature Review

Several researchers have already explored this field and many papers have been published. This section deals with the various practices already employed. This field has its base on natural language processing, machine learning and cognitive science. In [1] Pakistan Election results were predicted, they used quantitative behavior into account and suggested the top 5 parties. A detailed case study for Singapore Elections was also carried out using sentiment analysis in 2011. In [2] automatic buzzer detection and sentiment analysis was the topic. Here, buzzers were bots or paid users who tweeted in favor of a party for monetary gain. In [3] the data mining techniques were used and for polarity classification SentiWordNet 3.0.0 and concluded that k-nearest neighbor is better than Random Forest, Naive Bayesian, Bays Net. In [4] a detailed account of Delhi CM elections was carried out. Tweets on Arvind Kejriwal and Kiran Bedi were collected over a few weeks and the SentiWordNet and Word Sense Disambiguation were used for determining political polarity.

In [5] Vader was used to label the tweets and a two stage framework and entity classifier to predict results. In [6] there was a discussion about the different approaches for sentiment analysis – lexicon_based approach, machine_learning approach and a hybrid approach. Analyzing sentiments from tweets has been carried out by many scholars using R or in python (with advanced libraries scipy, scikit, nltk) or using intelligent APIs.

Table 1 shows the various trials by researchers using different approaches. The results vary because the training datasets are different, sources vary, domains differ. We will talk about these approaches, in the next few sections. The accuracy can vary minutely each time live data is collected. In our paper, we will deal with Naive Bayes (NB), Support Vector Machine(SVM) and Maximum Entropy(ME) classifiers.

3 Methodology

3.1 Data Collection

Firstly, our aim is to collect relevant tweets. This is ensured by the Twitter Streaming API, where a user can create a Twitter app and use the access keys and consumer tokens generated, for authentication. Once authenticated, we can fetch live tweets in

Table 1. Accuracy for various approaches used:

Paper by	Technique	Accuracy(%)
Ravikiran Janardhana [9]	Naive Bayes, Maximum Entropy, SVM	66.82, 60.35, 63.90
Jyoti Ramteke et al. [5]	nltk (SVM), Scikit-learn (SVM)	54.0, 99.0
Varghese S Chooralil, Rincy Jose [4]	SentiWordNet, Naive Bayes, HMM, classifier using ensemble approach	21.05, 69.92, 64.06, 71.48
Anurag P. Jain, Mr. Vijay D. Katkar [3]	k-nearest neighbour, RandomForest, BaysNet, NaiveBayes	96.6398, 65.6681, 48.9579, 60.3159
Farhan Hassan Khan, Usman Qamar [7]	ML and lexicon	85.7
Zhang et al. [8]	ML and lexicon	85.4

general or of a particular domain or area of interest. Fetching of tweets requires several parameters, language of text (say, English) or the fetching a specific number of tweets (say, 100). Once the tweets are collected we proceed with rest of the process.

3.2 Pre-processing of Data

The data that is collected cannot be used directly for predicting results. The data is unstructured and has to be processed before we proceed further. This step is crucial because it affects the overall accuracy of the results.

Using regular expressions we try to eliminate URLs, hashtags ‘#’, usernames addressed by ‘@’ and other such special characters and stopwords to avoid unnecessary details. Redundant tweets are also eliminated.

3.3 Training Dataset and Classifiers

Now the classifiers are trained using the clean data so that it is able to classify the tweets in future. Therefore, the training dataset should cover almost all domain specific tweets to avoid overfitting and underfitting and to ensure better results. For example, implementing SVM in python, off the 100%, 80% are used for training the classifier and the rest 20% for testing. Once the classifier is trained it can classify new tweets based on its prior knowledge it obtained while training. Now, when real time tweets are collected they are classified accordingly. This step is necessary for supervised methods. The classification approaches are dealt with greater depth in the next section.

4 Approaches for Sentiment Analysis

There are two main approaches for performing sentiment classification.

4.1 Lexicon-Based Approach

Lexicon means a dictionary, in our case a set of words with predefined polarity. In this method, the tweets after pre-processing are first split into smaller tokens, this is known as tokenization. Now our tokens are compared with the lexicons. Their sentiment polarity is determined positive, negative or neutral based on the collection of words in the lexicon which is like a storehouse of words with sentiment values assigned beforehand by the researchers. Textblob uses this approach to classify tweets. This is a simple approach, easy to implement but the results are not very accurate. This is an unsupervised method where no training is required.

4.2 Machine Learning Based Approach

This is a supervised learning method where data is trained and tested. Three most common approaches are given below:

- **Naive Bayes** (NB): NB is a model which works good on text classification. NB algorithm is used to classify textual data. NB has many advantages such as simple training method and lower complexity. But during text classification in the higher dimension, NB suffers greatly due to unavailability of data. This works on Bayes Theorem, where a tweet is classified based on the training dataset. This happens when our data set consist data like tweets and the training set is small because we are manual classifying each tweet.
- **Support Vector Machine** (SVM): SVM is a type of classification method which outputs an optimal hyperplane which classify our new example. In the literature survey, SVM performed efficiently for classifying our test data. SVM overcomes the problem of NB and is effective in high dimensional spaces. Also it is memory efficient because it uses a subset of training points called as Support Vector. Each vector has an entry corresponding to a feature. But if the number of tweets is very less as compare to the number of features, SVM is likely to give poor results.
- **Maximum Entropy** (ME): ME classifier follows exponential model for classification. Its assumes conditionally independent features and works on ME principle i.e. from all the models that fit our training dataset, the one with the largest entropy is selected where, largest entropy is the probability distribution which best represents the current state of knowledge. It is useful for sentiment analysis and text classification Fig. 1.

5 Analysis and Results

Based on the input sources our results show that in the current scenario, BJP has a greater support when compared to Congress, as its negative polarity is greater. The graph (Fig. 2) shows the plot for both the political parties. We used three classifiers namely, Naive Bayes, Support Vector Machine, Maximum Entropy.

Accuracy (A) is the ratio of correctly predicted tweets.
Precision (P) is the fraction of retrieved data that is relevant.

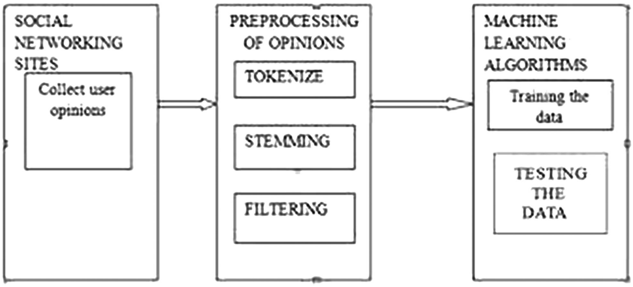


Fig. 1. Sentiment analysis methodology [11]

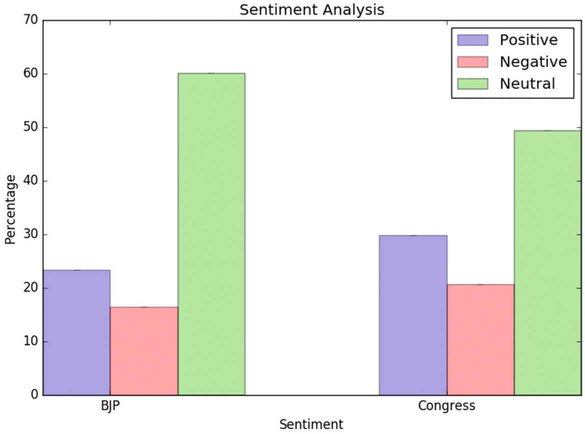
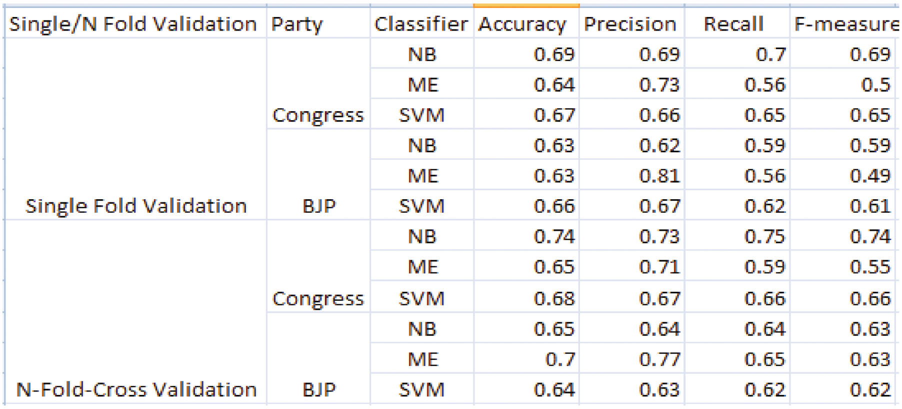
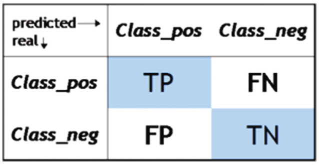


Fig. 2. Graph plot for BJP and Congress



Single/N Fold Validation	Party	Classifier	Accuracy	Precision	Recall	F-measure
Single Fold Validation	Congress	NB	0.69	0.69	0.7	0.69
		ME	0.64	0.73	0.56	0.5
		SVM	0.67	0.66	0.65	0.65
	BJP	NB	0.63	0.62	0.59	0.59
		ME	0.63	0.81	0.56	0.49
		SVM	0.66	0.67	0.62	0.61
N-Fold-Cross Validation	Congress	NB	0.74	0.73	0.75	0.74
		ME	0.65	0.71	0.59	0.55
		SVM	0.68	0.67	0.66	0.66
	BJP	NB	0.65	0.64	0.64	0.63
		ME	0.7	0.77	0.65	0.63
		SVM	0.64	0.63	0.62	0.62

Fig. 3. Results for single fold and N-fold cross validation



predicted→ real↓	*Class_pos*	*Class_neg*
Class_pos	TP	FN
Class_neg	FP	TN

Fig. 4. Classification matrix [10]

Recall (R) is the fraction of data that are relevant and successfully retrieved i.e. correctly predicted positive tweets.
F-measure (F) is the harmonic mean of precision and recall.

Figure 5 shows graphically the comparison of these values for the political parties used from the results as shown in Fig. 3.

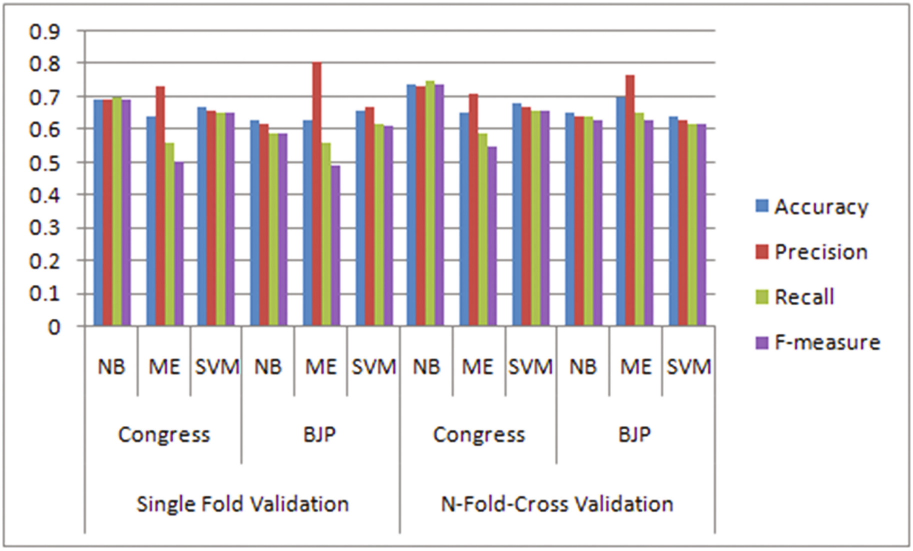


Fig. 5. Plot shows accuracy, precision, recall, F-measure values for single and n-cross validation

$$\text{Precision} = \text{tp}/(\text{tp} + \text{fp}). \tag{1}$$

$$\text{Recall} = \text{tp}/(\text{tp} + \text{fn}). \tag{2}$$

$$\text{F} = 2 * (\text{P} * \text{R})/(\text{P} + \text{R}). \tag{3}$$

These results may vary based on the real time data that comes in and also based on the training datasets. Currently the lowest accuracy classifier may give highest accuracy for another dataset. Results of such online data combined with data from surveys held offline can yield better solutions since rural areas don't use such social networking platforms (Fig. 4). The results for BJP and congress are as means_sentiment (BJP, Congress):

means_positive = (23.33%, 29.83%)
means_negative = (16.50%, 20.68%)
means_neutral = (60.16%, 49.47%)

6 Conclusion and Future Work

In this paper, we discussed about the sentiment analysis in political field. We began with a training dataset, collected live tweets and classified their sentiments using various classifiers. We used various machine learning models to classify the tweets into pre-defined categories – positive, negative and neutral.

Our concentration was focused on two political parties in India-"BJP" and "Congress". After the classification we compared their overall (positive, neutral and negative) polarity to predict which party is supported by the common masses and will have higher probability of winning the upcoming Elections. A graph was plotted for the results and with an accuracy of about 60–65% we are able to predict BJP's victory.

Also, we compared the results with the help of different classifiers. Automated classification makes it easier and faster to analyze data when compared to manual process which would consume significant amount of time and effort.

The sentiment classification approach can be extended to many fields like trend detection, popularity among the masses, dislikes or likes for a brand, customer reviewing. By using a sufficiently large and good training dataset accuracy of prediction can be improved. Innovative hybrid approaches could provide better results. This analysis can also be extended to multilingual data. There is need for good datasets in a wide range of time, since twitter limits the access to tweets over long periods of time. As people's opinions change over a span of time, election results need to monitor these changes as well to predict better results. Another extension could be employing deep learning techniques with neural networks which could handle sarcastic remarks and complicated negated statements better. By providing learning mechanism, and back-propagation accuracy is bound to improve.

We can also have a game theory approach based on emotional evolution prediction algorithm which combines the affective computing, in which the mixed nash equilibrium strategies are calculated as the future emotional behaviour of interactive users [12].

References

1. Razzaq, M.A., Qamar, A.M., Bilal, H.S.M.: Prediction and analysis of pakistan election 2013 based on sentiment analysis. In: 2014 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (2014)

2. Ibrahim, M., Abdillah, O., Wicaksono, A.F., Adriani, M.: Buzzer detection and sentiment analysis for predicting presidential election results in twitter nation. In: 2015 IEEE 15th International Conference on Data Mining Workshops (2015)
3. Jain, A.P., Katkar, V.D.: Sentiments analysis of twitter data using data mining. In: 2015 International Conference on Information Processing (ICIP) Vishwakarma Institute of Technology, December 2015
4. Rincy, J., Varghese, S.C.: Prediction of Election Result by Enhanced Sentiment Analysis on Twitter Data using Classifier Ensemble Approach (unpublished)
5. Jyoti, R., Darshan, G., Samarth, S., Aadil, S.: Election Result Prediction Using Twitter sentiment Analysis (unpublished)
6. Kharche, S.R., Bijole, L.: Review on sentiment analysis of twitter data. Int. J. Comput. Sci. Appl. **8**, 53–56 (2015)
7. Khan, F.H.: TOM: twitter opinion mining framework using Hybrid Classification scheme. Decision Support Syst. **57**, 245–257 (2014)
8. Zhang, L., Ghosh, R., Dekhil, M., Hsu, M., Liu, B.: Combining lexicon-based and learning-based methods for twitter sentiment analysis. Technical report, HP Laboratories (2011)
9. Ravikiran, J.: Twitter Sentiment Analysis and Opinion Mining (unpublished)
10. http://algolytics.com/tutorial-how-to-determine-the-quality-and-correctness-of-classification-models-part-1-introduction/
11. http://article.sciencepublishinggroup.com/html/10.11648.j.ijdst.20160204.11.html
12. http://dl.acm.org/citation.cfm?id=2936576

An Efficient Method for Detecting Electrical Spark and Fire Flame from Real Time Video

Sonia Corraya and Jia Uddin(✉)

Department of Computer Science and Engineering,
BRAC University, Dhaka, Bangladesh
{sonia.corraya, jia.uddin}@bracu.ac.bd

Abstract. Prompt fire detection and localization is an essential requirement for saving lives and reducing damages caused by fire accidents. The main source of 39–45% fire accidents is electrical origin. As electric fire accidents are increasing, detecting electrical spark and fire flame from the common origination point-electrical socket is imperative and vital. In this paper an efficient method for detecting electrical socket spark and fire flame from real time video processing is proposed. At first, extracted image frames from video are converted from RGB to YCbCr. After that, any change in average Luminance Y, average Blue component Cb, average Red component Cr triggers moving foreground detection step. From the detected moving foreground by frame differencing, regions having highest Luminance are considered as suspects. Finally, the actual spark or fire flame regions are detected by taking only those suspects whose area changes in consecutive frames. For evaluating the proposed method, two types of dataset (electric spark and fire flame) are used. Experimental result shows 80% and 100% accuracy in spark and flame detection respectively. The proposed algorithm worked properly in the five tests for flame and in the four tests for spark. In addition, we have compared the performance of proposed method with a state of art model and experimental results show that the proposed model outperforms existing system with 60% more accuracy. The proposed system can assist in fire accident investigations and in proper prevention action decision making by early fire and fire source detection.

Keywords: Electrical fire · Socket · Spark · YCbCr · Fire · Flame

1 Introduction

From small buildings to large towers there is always risk of violent electrical fire incident from a single spark or fire flame. Electrical origin is one of major cause of fire accidents and the number is increasing with the vast use of electricity. It is found that around 65% of fire deaths occur from fires because of the absence of any workable fire detection and extinguisher system [1]. Most of fire accidents start with electrical fire. According to [2], the main cause of 39–45% fire accidents is electrical origin.

Traditional systems detect fire after its origination and growing by burning objects, smoke creation, heat generation etc. [3–5]. Existing image processing based fire detection techniques considered fire color, temporal feature, smoke and region growing but till now no technique has considered detecting fire at the source. Rather than

S.M. Thampi et al. (eds.), *Advances in Signal Processing and Intelligent Recognition Systems*, Advances in Intelligent Systems and Computing 678,
DOI 10.1007/978-3-319-67934-1_32

waiting for fire to grow, to create smoke and to increase temperature, it is important to detect fire at the origination point. Video cameras used for surveillance service rarely cover electrical socket points hence failed to detect and localize electrical spark or fire flame at electrical plugs or sockets. This situation demands an efficient, fast and reliable fire detection method.

In this paper, an YCbCr color model based system is proposed for early electric spark and fire flame detection from electric origin point. The objective of this work is to detect fire as soon as it starts at electrical socket, plug, and cable points. General overview of the proposed system is shown in Fig. 1. The proposed system converts RGB image to YCbCr and detect Y, Cb, Cr changes in extracted video frame sequence. Moving pixels are identified by frame difference and connected moving regions having the maximum Luminance Y are detected as spark, fire flame at electric plug, socket points.

The primary reason for selecting YCbCr color space for this work is its ability to distinguish Luminance from Chrominance information [3]. Also, best result is found after considering Y rather than Cb or Cr of moving pixels. In [12] HSV and YCbCr both color models are used for flame color detection from background. For this work, brightness with the YCbCr color model is considered over pure color feature for two reasons, first is, spark is usually only of very bright white color and second is, gray scale image processing facilitates faster system response and cost effective system development. Brightest moving pixel area detection is a basic step of this system and YCbCr model is a good option for brightness related information extraction [8, 10, 11]. Two particular fire natures are considered in this research work: (i) electrical sparks do not change its center position but for its flickering characteristic consecutive frames have frame difference and moving pixels can be detected; and (ii) small fire flames in electrical plugs and wires do not grow quickly but move with almost same flame size.

From the related tasks, it can be seen that most of the systems have used RGB, HSV, La*b*, YCbCr color model with manual thresholding and smoke, temperature detector as the supporting system. Those systems take longer time to process and can only produce result after fire growing, smoke generating, heat increase and so on. Considering the above points, we proposed a single color model, YCbCr based system for electrical fire (spark and flame) detection that needs particularly less processing time and detect fire incident at its origin.

Exiting video sequence based fire detection methods can be classified in two classes: fire flame detection and smoke detection [12]. However to the best of our knowledge none of these researches have ever considered electrical sparks and fire sources. All of these approaches described below are for detecting fire behavior in specified environment (like forest, indoor, building). In [6] by using an adaptive threshold technique Hong and Wenhao able to extract the particulars flame objects. They also used flame color in order to analysis the detect fire. In [8], a different mechanism is used for separating Luminance from Chrominance by YCbCr colors space. As we know noise distortion is a great effect on fire detection. So, we observed that, in [9] median filters are used for decreasing noise and parallel for fire detection using Bayes classifier. In [10], a system is described for forest fire detection that is also focused on YCbCr color space. These types of color space are also used in [11]. We can find that, in [7], Chen et al. make a combination of RGB based color space with

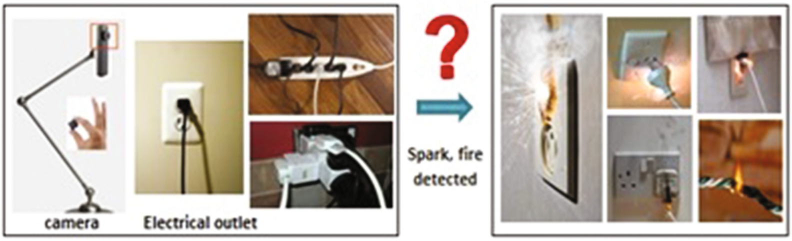


Fig. 1. A general overview of electrical spark or fire detection method

saturation to make a decision formula about the potentiality of fire pixels. A general overview of an electrical spark or fire detection system is shown in Fig. 1.

The rest of the paper is organized as follows: in Sect. 2, the proposed methodology is presented and described step by step in detail. Experimental results and discussion are presented in Sect. 3 and finally, conclusions are drawn in Sect. 4.

2 Proposed System

The proposed system is described by five steps as shown in Fig. 2 and pseudo code of the proposed model in given in Fig. 4.

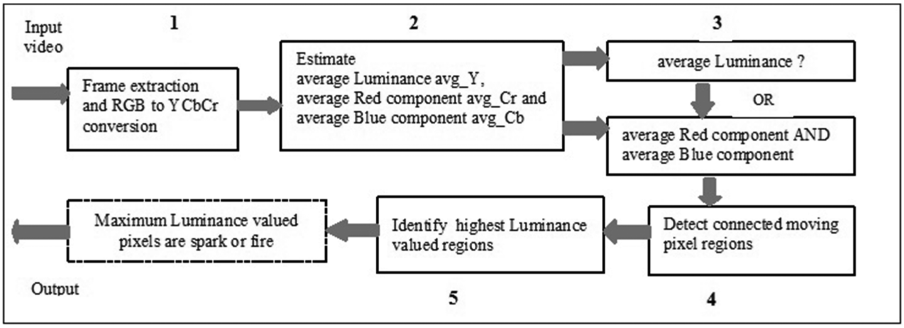
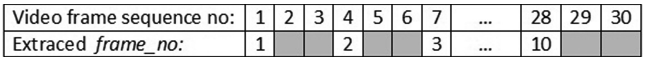

Input video
1
Frame extraction and RGB to YCbCr conversion
2
Estimate average Luminance avg_Y, average Red component avg_Cr and average Blue component avg_Cb
3
average Luminance ?
OR
average Red component AND average Blue component
Detect connected moving pixel regions
4
Identify highest Luminance valued regions
5
Maximum Luminance valued pixels are spark or fire
Output

Fig. 2. Steps of the proposed system

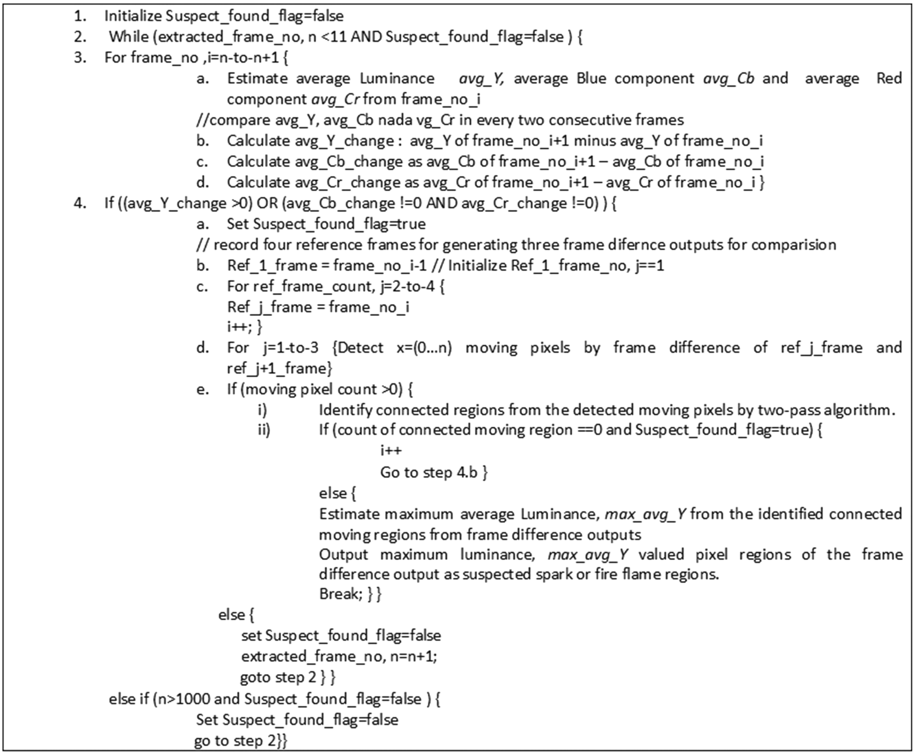
Step 1: Frame extraction and RGB to YCbCr conversion

From a 30 frame per second input video sequence only 10 frames are extracted and converted from RGB to YCbCr. As shown in Fig. 3, after the first frame of the video sequence, two frames are skipped.



Video frame sequence no:	1	2	3	4	5	6	7	...	28	29	30
Extraced *frame_no:*	1			2			3	...	10		

Fig. 3. Frame extraction form input video



```
1.  Initialize Suspect_found_flag=false
2.   While (extracted_frame_no, n <11 AND Suspect_found_flag=false ) {
3.  For frame_no ,i=n-to-n+1 {
            a.  Estimate average Luminance  avg_Y, average Blue component avg_Cb and  average  Red
                component avg_Cr from frame_no_i
            //compare avg_Y, avg_Cb nada vg_Cr in every two consecutive frames
            b.  Calculate avg_Y_change : avg_Y of frame_no_i+1 minus avg_Y of frame_no_i
            c.  Calculate avg_Cb_change as avg_Cb of frame_no_i+1 – avg_Cb of frame_no_i
            d.  Calculate avg_Cr_change as avg_Cr of frame_no_i+1 – avg_Cr of frame_no_i }
4.  If ((avg_Y_change >0) OR (avg_Cb_change !=0 AND avg_Cr_change !=0) ) {
            a.  Set Suspect_found_flag=true
            // record four reference frames for generating three frame difernce outputs for comparision
            b.  Ref_1_frame = frame_no_i-1 // Initialize Ref_1_frame_no, j==1
            c.  For ref_frame_count, j=2-to-4 {
                Ref_j_frame = frame_no_i
                i++; }
            d.  For  j=1-to-3  {Detect  x=(0…n)  moving  pixels  by  frame  difference  of  ref_j_frame  and
                ref_j+1_frame}
            e.  If (moving pixel count >0) {
                    i)      Identify connected regions from the detected moving pixels by two-pass algorithm.
                    ii)     If (count of connected moving region ==0 and Suspect_found_flag=true) {
                                i++
                                Go to step 4.b }
                            else {
                            Estimate maximum average Luminance, max_avg_Y from the identified connected
                            moving regions from frame difference outputs
                            Output maximum luminance, max_avg_Y valued pixel regions of the frame
                            difference output as suspected spark or fire flame regions.
                            Break; } }
               else {
                  set Suspect_found_flag=false
                  extracted_frame_no, n=n+1;
                  goto step 2 } }
    else if (n>1000 and Suspect_found_flag=false ) {
            Set Suspect_found_flag=false
            go to step 2}}
```

Fig. 4. Pseudo code of proposed method

Step 2: Estimation of average Luminance Y, average Red component Cr and average Blue component Cb

From the first extracted frame, average Luminance Y as avg_Y, average Red component Cr as avg_Cr and average Blue component Cb as avg_Cb are estimated.

Step 3: Check for change in average Luminance, average Red component, average Blue component

All three components avg_Y, avg_Cr and avg_Cb are compared in every two consecutive frames. If any change is detected in avg_Y or any change in avg_Cr and avg_Cb then initial decision is made that suspect spark or fire flame is found. For confirming and localizing the suspect(s) found in frame_i, next two consecutives frames including frame_i and frame_i-1 are recorded as reference frames as shown in Fig. 5(a)–(d). These four reference frames are investigated in next step.

Step 4: Detect connected moving pixel regions

For any suspect found in previous step, moving pixels are detected by frame differencing over consecutive reference frames. If any moving pixel found, then two pass algorithm [18] is applied for detecting connected moving regions as shown in Fig. 5(e)–(g).

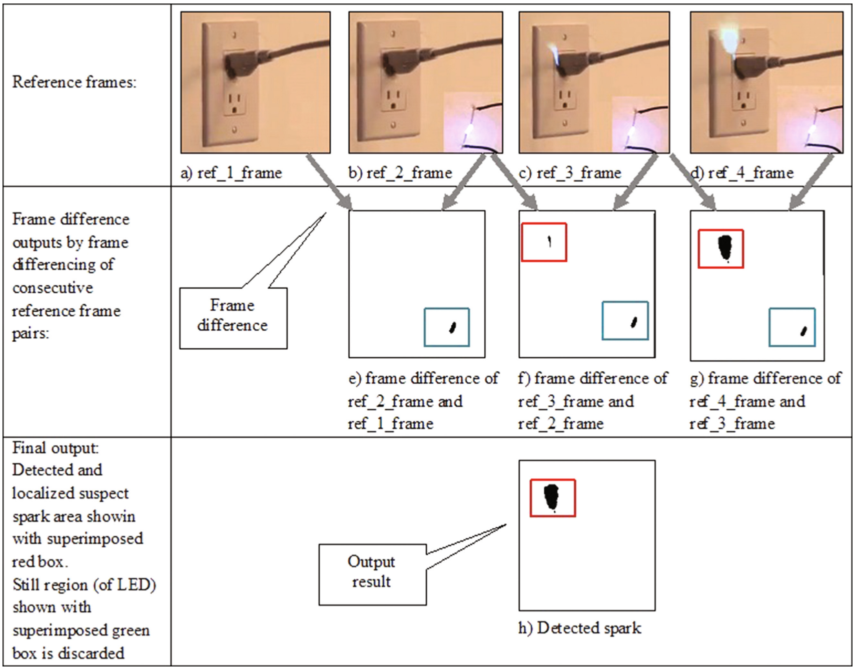

Reference frames:
a) ref_1_frame
b) ref_2_frame
c) ref_3_frame
d) ref_4_frame
Frame difference outputs by frame differencing of consecutive reference frame pairs:
Frame difference
e) frame difference of ref_2_frame and ref_1_frame
f) frame difference of ref_3_frame and ref_2_frame
g) frame difference of ref_4_frame and ref_3_frame
Final output: Detected and localized suspect spark area showin with superimposed red box. Still region (of LED) shown with superimposed green box is discarded
Output result
h) Detected spark

Fig. 5. (a)–(d). Video frame sequence [19]; (e)–(g) frame differences and (h) final output

Step 5: Identify highest Luminance valued regions from frame difference outputs

From corresponding reference frame, identify the highest Luminance value of the connected moving regions that are detected from the frame difference outputs. Maximum Luminance value is considered here for ignoring any other moving objects. As shown in Fig. 5(h), still region shown with green box are discarded and only changing or moving regions, shown with red box are considered as suspected spark or fire flame area.

3 Results and Discussion

As no electrical fire dataset is found in the internet and making spark for real in home environment is risky, spark and fire flame video files are collected from internet [21, 22] and extracted frame is modified accordingly for testing various scenarios as

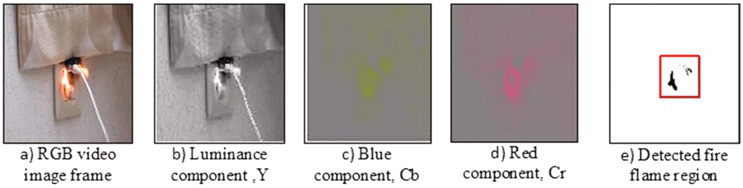


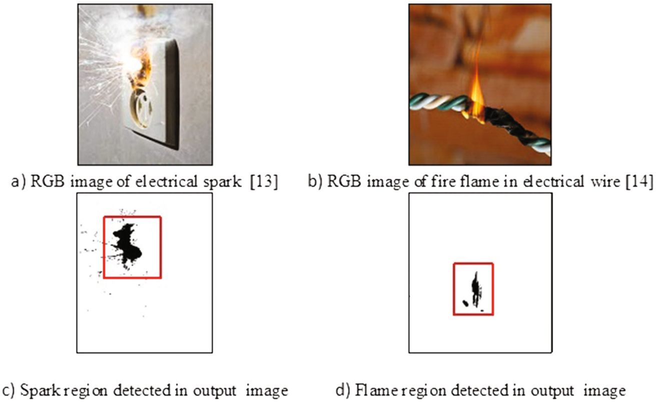
Fig. 6. (a)–(e). System output of fire flame [20]

shown in Fig. 5. In Fig. 5, still LED of similar brightness of the spark is added to check if it is discarded in the final result after applying the proposed model. An example of three different components (Y, Cb and Cr) of a single frame is shown in Fig. 6.

For evaluating the performance of the proposed model, five spark video and five flame video are selected. As shown in Table 1 and in Fig. 8, experimental result shows 80% accuracy in electric spark detection and 100% accuracy in fire flame detection. Accuracy of flame detection is higher because flame usually presents in all consecutive image frame whereas spark might be visible only once in a single video. Such spark scenarios are not considered as serious spark incident. False positive error can be occurred if there is any moving LED or light bulb but this case is quite uncommon especially in electrical socket area. False negative error can be occurred if any spark or fire flame goes outside of capturing video range immediately after the origination. Figure 7 shows some example of the system output.

Table 1. Accuracy rate of the proposed system

Experimental data type	Total number of case/video	No. of successful detection cases	Accuracy rate (%)
Electrical spark	5	4	(4/5) * 100 = 80
Fire flame	5	5	(5/5) * 100 = 100


a) RGB image of electrical spark [13]

b) RGB image of fire flame in electrical wire [14]

c) Spark region detected in output image

d) Flame region detected in output image

Fig. 7. Example of the proposed system output result

Though this is the very first method of this kind, an existing image processing based fire detection approach [12] is used for comparative evaluation. The comparison results are presented in Tables 1, 3 and 4. Table 2 and Fig. 9 shows that proposed method is 60% more accurate than the traditional method for electrical fire detection.

Proposed method can detect spark and fire flame faster than traditional method and detail time frame (time-stamp of frame) comparison results are presented in Tables 3 and 4. Figure 10 shows the time frame comparison for both data types graphically.

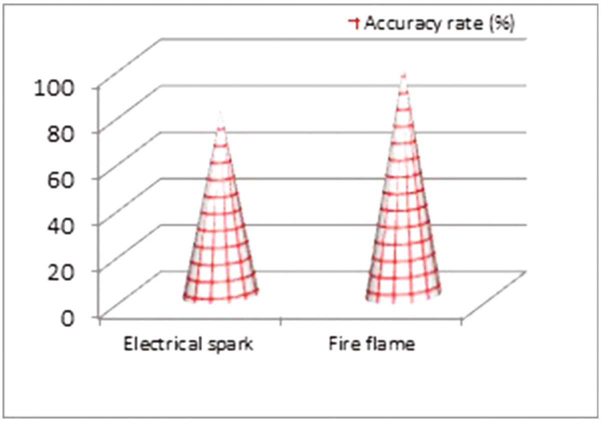


Fig. 8. Accuracy rate of the proposed method for electrical spark and fire flame

Table 2. Comparison result of detection accuracy

Experimental data type	Detected by proposed model (out of 5 cases)	Detected by traditional model (out of 5 cases)	Accuracy rate of the proposed model (%)	Accuracy rate of the traditional method (%)
Electrical spark	4	1	(4/5) * 100 = 80%	(1/5) * 100 = 20%
Fire flame	5	2	(5/5) * 100 = 100%	(2/5) * 100 = 40%
			Average : 90%	Average: 30%

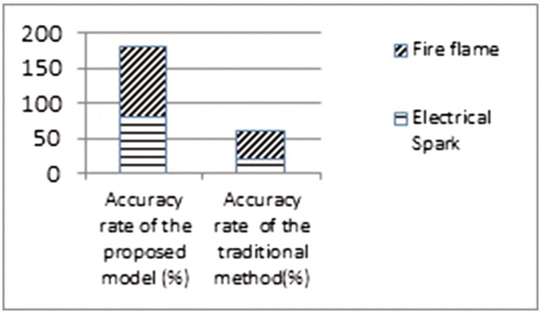


Fig. 9. Comparison of accuracy rate between the proposed model and traditional method

Table 3. Comparison result of detection time frame for electrical spark

Electrical spark video/scenario no.	Detection time frame by proposed model (s)	Detection time frame by traditional method (s)
1	1.05	Not detected
2	3.0142	14.033 s
3	4.01	Not detected
4	Not detected	Not detected
5	4.0166	Not detected
Average:	3.0227 s	14.33 s

Table 4. Comparison result of detection time frame for fire flame

Fire flame video/scenario no.	Detection time frame by proposed model (s)	Detection time frame by traditional method (s)
1	8.0125	Not detected
2	6.01428	Not detected
3	7.1	130 s
4	9.0333	Not detected
5	10.02	80.0166 s
Average :	8.036016 s	105.0083 s

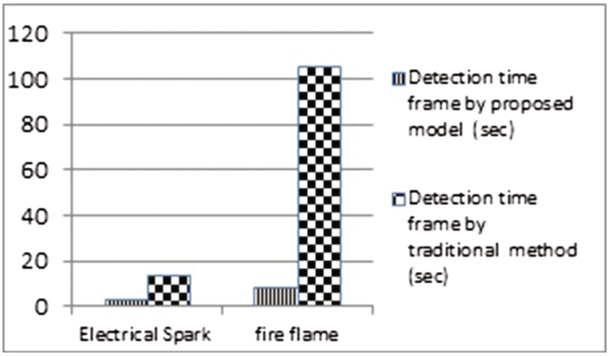


Fig. 10. Comparison of detection time frame between the proposed model and traditional method

4 Conclusions

In this paper a new YCbCr model based systems is proposed which covers electrical socket area with camcorder and detect any spark and fire flame in the electrical socket, plug and cables. Image frames are extracted from video sequence and then frame difference is used to check for spark and fire flame suspects by spotting moving pixels in consecutive frames. After that, two-pass algorithm is used for connected region detection from identified moving pixels for detecting suspect growths and movements.

Both, experimental result with two types of dataset (electric spark and fire flame) and comparison result indicate promising system performance. In future, we will extend this system to cover all kinds of outdoor electrical fire origin. The proposed system will warn faster and can save from big accidents. Also, this system can be used in fire accident investigations and will assist in deciding proper fire prevention action as different fire accident type need different prevention method.

References

1. ESFI: Home electrical fires facts, statistics and safety tips. http://www.esfi.org/resource/home-electrical-fires-184. Accessed 19 Nov 2016
2. Buildings not inspected for electrical safety – Andhra Pradesh – The Hindu. http://www.thehindu.com/todays-paper/tp-national/tp-andhrapradesh/buildings-not-inspected-for-electrical-safety/article3740329.ece. Accessed 19 Nov 2016
3. Li, M., et al.: Review of fire detection technologies based on video image. J. Theor. Appl. Inf. Technol. **49**(2) (2013)
4. Habiboğlu, Y.H., Günay, O., Çetin, A.E.: Covariance matrix-based fire and flame detection method in video. Mach. Vis. Appl. **23**(6), 1103–1113 (2012)
5. Dimitropoulos, K., Tsalakanidou, F., Grammalidis, N.:. Flame detection for video-based early fire warning systems and 3D visualization of fire propagation. In: 13th IASTED International Conference on Computer Graphics and Imaging (CGIM 2012), Crete, Greece (2012)
6. Wang, W., Zhou, H.: Fire detection based on flame color and area. In: 2012 IEEE International Conference on Computer Science and Automation Engineering (CSAE), vol. 3. IEEE (2012)
7. Chen, J., He, Y., Wang, J.: Multi-feature fusion based fast video flame detection. Build. Environ. **45**(5), 1113–1122 (2010)
8. Celik, Turgay, Demirel, Hasan: Fire detection in video sequences using a generic color model. Fire Saf. J. **44**(2), 147–158 (2009)
9. Lei, W., Liu, J.: Early fire detection in coalmine based on video processing. In: Proceedings of the 2012 International Conference on Communication, Electronics and Automation Engineering. Springer, Heidelberg (2013)
10. Celik, T., Demirel, H.: Fire detection in video sequences using a generic color model. Fire Saf. J. **44**(2), 147–158 (2009)
11. Liu, Z.-G., Zhang, X.-Y., Wu, C.-C.: A flame detection algorithm based on Bag-of-Features in the YUV color space. In: 2014 International Conference on Intelligent Computing and Internet of Things (ICIT). IEEE (2015)
12. Seebamrungsat, J., Praising, S., Riyamongkol, P.: Fire detection in the buildings using image processing. In: 2014 Third ICT International Student Project Conference (ICT-ISPC). IEEE (2014)
13. What are electrical accidents ? A blog on electricity. http://onelectricity.blogspot.com/2014/08/what-are-electrical-accidents.html. Accessed 23 Nov 2016
14. Short circuit, burnt cable, on dark color background stock photo 178505459 : shutterstock. http://www.shutterstock.com/pic-178505459/stock-photo-short-circuit-burnt-cable-on-dark-color-background.html?src=PbKQC-HrKKVxJEQlIEA2Mw-1-2. Accessed 23 Nov 2016
15. Fire safety tips: what to do in a fire. http://www.healthxchange.com.sg/healthyliving/HealthatWork/Pages/Fire-Safety-Tips-What-to-Do-in-a-Fire.aspx. Accessed 23 Nov 2016

16. Dangerous wall plug stock photo, royalty free image: 632848 – Alamy. http://www.alamy.com/stock-photo-dangerous-wall-plug-632848.html. Accessed 23 Nov 2016
17. Electrical safety for diy electronic circuits – Electronics information from PenguinTutor. http://www.penguintutor.com/electronics/electrical-safety. Accessed 23 Nov 2016
18. Connected-component labeling – Wikipedia. https://en.wikipedia.org/wiki/Connected-component_labeling. Accessed 19 Nov 2016
19. Plug-In-connecting-Sparks.mov – YouTube. https://www.youtube.com/watch?v=3bbjb8W2al8. Accessed 23 Nov 2016
20. ダイソーのコンセントカバーが便利。人気商品と魅力を紹介 | iemo[イエモ]. https://iemo.jp/47124. Accessed 23 Nov 2016
21. 1mp H.264 1280x720 wireless hidden mini ip camera hd wifi, View mini ip camera hd wifi, Hichip or OEM Product Details from Shenzhen Hichip Vision Technology Co., Ltd. on Alibaba.com. https://hichip.en.alibaba.com/product/60475676491-800790310/1mp_H_264_1280x720_wireless_hidden_mini_ip_camera_hd_wifi.html. Accessed 19 Nov 2016
22. High voltage experiments – Homemade stun guns and crazy Tesla FAILS – Joe Genius – YouTube. https://www.youtube.com/watch?v=L2F8kRvjhTY. Accessed 19 Nov 2016

Speaker-Independent Automatic Speech Recognition System for Mobile Phone Applications in Punjabi

Puneet Mittal[1](✉) and Navdeep Singh[2]

[1] BBSB Engineering College, Fatehgarh Sahib, India
puneet.mittal@bbsbec.ac.in
[2] Mata Gujri College, Fatehgarh Sahib, India
navdeep_jaggi@yahoo.com

Abstract. Speaker-independent Automatic Speech Recognition (ASR) system based mobile phone applications are gaining popularity due to technological advancements and accessibility. Speech based applications may provide mobile phone accessibility and comfort to people performing activities where hand-free phone access is desirable e.g. drivers, athletes, machine operators etc. Similarly, users with disabilities like low vision, blindness and physically challenged may use it as an assistive technology. Development of ASR system for a specific language needs accurate, reliable and efficient acoustic model having language-specific pronunciation dictionary. Punjabi language is one of the popular languages worldwide having more than 150 million speakers. Three acoustic models- continuous, semi-continuous and phonetically-tied are developed based on three pronunciation dictionaries- word, sub-word and character based. Analysis of performance results validate Punjabi language principle "One word one sound" by having better accuracy and reliability for character based pronunciation dictionary than others. Further, phonetically-tied model outperforms others in terms of accuracy, word error rate and size due to reasonable number of Gaussians.

Keywords: Acoustic model · Language model · Punjabi · CMU sphinx · Dictionary

1 Introduction

Automatic speech recognition (ASR) is the transcription of speech signal into readable text to identify and process human voice. Speech provides vocalized communication through large vocabularies having different words formed out of phonetic combination of sound units called phoneme. Based upon the vocabulary size, a word may have phonetic representation as a word itself for small vocabulary, syllable-based or sub-word based representation for large vocabulary, or character-based representation for languages having characters with distinct sound. ASR systems are being widely used in various applications for desktop, laptop and hand-held devices like mobile phones, where each application has its own set of requirements. High speed is desirable for real-time applications while accuracy is the key concern for command and control

S.M. Thampi et al. (eds.), *Advances in Signal Processing and Intelligent Recognition Systems*, Advances in Intelligent Systems and Computing 678,
DOI 10.1007/978-3-319-67934-1_33

applications and dictation applications. Efficient space utilization and high speed is desirable for mobile phone applications while high speed is expected from desktop applications having ample storage space available.

ASR systems have been an active area of research for almost last six decades. Research started in this field in the year 1952 with the development of Aurdey [4], a speaker-dependent speech recognizer having 97–99% digit recognition accuracy. It was followed by DoD's DARPA Speech Understanding Research (SUR) program [10] and Carnegie Mellon's "Harpy" speech-understanding system [16] having ability to recognize 1011 words. Hidden Markov Model (HMM) based methods gained popularity in 1980s and are still being widely used.

A major revolution in this field came in the year 1990 with the development of Sphinx [15]. Sphinx is an accurate, large vocabulary, speaker independent, continuous speech recognition system. It introduced three acoustic models- continuous [19], semi-continuous [7] and phonetically-tied [6]. They differ in the way their mixture of Gaussians is built, that is used to compute the score of each frame. In continuous model every senone has its own set of gaussians thus the total number of gaussians in the model is about 150 thousand. It requires much processing to compute the mixture efficiently. In semi-continuous model, there are total 700 gaussians for use with different mixtures to score the frame. Due to the smaller number of gaussians semi-continuous models are fast, but because of more hardcoded structure their accuracy is low as compared to continuous models. Phonetically-tied models (PTM) use about 5000 gaussians thus providing better accuracy than semi-continuous. It achieves almost same accuracy as of continuous model with less processing and storage requirements. So, it is significantly faster than continuous models and can be used for mobile applications.

ASR system requires development of an efficient acoustic model based on language specific pronunciation dictionary. This paper proposes the development of efficient acoustic model for Punjabi language that can be used to build ASR system for mobile phone applications. Section 2 covers the related work in the field of ASR followed by problem formulation in Sect. 3. Section 4 gives introduction to Punjabi language while proposed methodology is explained in Sect. 5. Section 6 gives detailed development of ASR system for Punjabi language. Results are analyzed in Sect. 7 followed by discussion and Conclusions are given in Sect. 8.

2 Related Work

Various ASR systems have been proposed by researchers from time to time. Most of the applications like Google Voice Search [23] are in English, Spanish or other European languages. Wang et al. [27] developed ASR for Chinese having complete recognition of continuous Mandarin speech with large vocabulary. Walha et al. [26] developed ASR for Standard Arabic language using HTK toolkit. Satori et al. [22] trained a model for Amazigh using CMU Sphinx tools having 92.89% accuracy for 16 GMM. Naing et al. [18] developed large vocabulary continuous speech recognition system for Myanmar language using deep neural network approach. Researchers are also working on other Asian languages like Japanese, Korean [24] etc. Indian languages like Hindi [11], Assamese [1], Tamil [25], Bengali [3] etc. are also being explored. Till now, little work has been done on speech recognition in Punjabi [12–14]. Dua et al. [5] proposed isolated word ASR system for Punjabi using HTK toolkit with overall system performance of 95.63% for a limited vocabulary having 115 Punjabi words.

3 Problem Formulation

Mobile phones have become future communication instruments by replacing computers and laptops with the advent of better hardware, computation and storage capabilities, and battery technology improvements. Speaker-dependent applications embedded in mobile phones are being ignored by the majority of users due to usability, accuracy, speed and storage constraints. Speaker-independent applications as a low cost, high capacity alternative to speaker-dependent applications are being developed to provide user-friendly, accurate, fast and low memory interface [17] for simple features like phone dialing and dictation to complex command and control features. It covers speech based applications like continuous digit dialing, name dialing, command and control for menus and navigation systems, games, and interactive man-machine interfaces.

Speech based mobile phone applications provide accessibility and comfort in situations where a person is driving a vehicle or doing some activity and needs to dial a phone number, send SMS or use GPS etc. It acts as an assistive technology for users with disabilities like low vision, blindness and physically challenged.

Punjabi is the native language of people of Punjab state in India. It is spoken by more than 150 million native speakers worldwide. According to a report by the Commissioner for Linguistic Minorities [2], 91.69% people speak Punjabi in Punjab state. 62.52% people of Punjab live in rural area [20]. People from rural areas of Punjab cannot use speech based applications built in foreign languages. So there is a need to develop speaker-independent Punjabi based applications for mobile phones. Currently, there is no acoustic model specifically built for mobile phone applications in Punjabi. This paper aims to build efficient acoustic model which can be used to develop speaker-independent mobile phone applications for Punjabi.

4 Punjabi Language

Punjabi is an Indo-Aryan language [21] widely spoken in countries like India, Pakistan, Canada and UK. It is spoken by more than 150 million native speakers worldwide. Gurmukhi and Shahmukhi scripts are used for Punjabi in India and Pakistan respectively. Gurmukhi script being alphasyllabary in nature consists of two types of symbols- consonants and vowels. It is written from left-to-right and is spelled phonetically. Gurmukhi script is based on “one sound one symbol” principle.

Punjabi is a meaningful collection of sentences made up of words where each word is a collection of phones [9]. Punjabi is formed based on phones or sounds having 41 alphabets and 9 dependent vowels of Gurmukhi script. Out of 41 alphabets, 38 are consonants (from ਸ to ੜ਼) while 3 alphabets (ੳ, ਅ, ੲ) are used in independent vowel form. In addition to these, 3 auxiliary signs are also available as shown in Table 1. Words in Punjabi are formed from different combinations of consonants and dependent vowels

Table 1. Punjabi character set.

Consonants	-	-	-	ਸ	ਹ					
	ਕ	ਖ	ਗ	ਘ	ਙ					
	ਚ	ਛ	ਜ	ਝ	ਞ					
	ਟ	ਠ	ਡ	ਢ	ਣ					
	ਤ	ਥ	ਦ	ਧ	ਨ					
	ਪ	ਫ	ਬ	ਭ	ਮ					
	ਯ	ਰ	ਲ	ਵ	ੜ					
	ਸ਼	ਖ਼	ਗ਼	ਜ਼	ਫ਼	ਲ਼				
Independent vowels	ਅ	ਆ	ਇ	ਈ	ਉ	ਊ	ਏ	ਐ	ਓ	ਔ
Dependent Vowels		◌ਾ	ਿ◌	◌ੀ	◌ੁ	◌ੂ	◌ੇ	◌ੈ	◌ੋ	◌ੌ
Auxiliary Sign	◌ਂ	◌ੰ	◌ੱ							

like ਕਾ(ka), ਕਿ(Ki), ਕੀ(Ke), ਕੁ(Ku), ਕੂ(Koo), ਕੇ(Kae), ਕੈ (Kai), ਕੋ(Ko), ਕੌ(Kau). For example, the word ਚਾਰ is a combination of consonants ਚ and ਰ with vowel ◌ਾ forming ਚ ◌ਾ ਰ (CVC). A word may have a vowel at the beginning followed by one or more consonants and vowels e.g. ਇੱਕ. The words are joined together to form sentences as per language rules to make the sentences meaningful. In this paper Punjabi character set having 38 consonants, 10 independent vowels, 9 dependent vowels and 3 auxiliary signs is considered.

5 Proposed Methodology

Three methodologies are proposed for Punjabi phonetic representation based on words, sub-words and characters. In word-based methodology, each word is uniquely identified as an acoustic unit. As no segmentation of word into sub-words or characters is done, the word ਇੱਕ is represented as ਇੱਕ and ਤਿੰਨ as ਤਿੰਨ. In sub-word-based methodology, all characters of each word are scanned for identification of characters like consonants, dependent vowels, independent vowels and auxiliary signs to form sub-words based upon certain rules (Table 2). The word ਇੱਕ is segmented into two sub-words ਇੱ and ਕ while word ਤਿੰਨis segmented into sub-words ਤਿੰ and ਨ. In character-based methodology, each word is segmented into individual characters based upon certain rules (Table 3) and each character is stored in an array as a unique acoustic unit. The word ਇੱਕ is segmented into three characters ਇ, ◌ੱ and ਕ while word ਤਿੰਨ is segmented into four characters ਤ, ਿ◌, ◌ੰ and ਨ.

Further, three acoustic models- continuous, semi-continuous and phonetically-tied are developed for words, sub-words and characters identified with the above three methodologies at different Gaussian densities (4, 8, 16, 32, 64, 128, 256). Detailed comparative analysis of these three acoustic models will be conducted for different

Table 2. Rules for Sub-word based segmentation.

Description	Pattern	Example of sub word
Independent Vowel	IV	ਇ (ਇ-IV)
Independent vowel followed by auxiliary sign	IV-AS	ਇੱ (ਇ –IV ◌ੱ -AS)
Consonant	C	ਕ (ਕ-C)
Consonant followed by Dependent Vowel	C-DV	ਦੋ (ਦ-C ◌ੋ-DV)
Consonant followed by Dependent Vowel and auxiliary sign	C-DV-AS	ਤਿੰ (ਤ-C ਿ◌ -DV ◌ੰ-AS)
Consonant followed by auxiliary sign	C-AS	ਪੰ (ਪ-C ◌ੰ-AS)

Table 3. Rules for character based segmentation.

Description	Pattern	Example
Independent Vowel	IV	ਇ
Consonant	C	ਕ
Dependent Vowel	DV	◌ੋ
Auxiliary Sign	AS	◌ੰ

performance parameters like Word Error Rate (WER), Accuracy, Speed, Size and Time taken to build the model. Based upon the outcome of comparative analysis, an optimal acoustic model will be recommended for the development of Automatic Speech Recognition model for Punjabi Mobile Applications.

6 Punjabi Speech Recognition System

This section describes the process of design and development of an efficient acoustic model for Punjabi automatic speech recognition system for mobile phone applications. Figure 1 shows the components of the proposed system. Initially the input speech signal is pre-processed at front end followed by extraction of acoustic features. The acoustic model, language model and dictionary are developed for the Punjabi, which are used by the speech recognition engine to identify the words spoken by the user. Speech corpus and Text corpus are the prerequisite for the development of Acoustic model and Language model respectively. The ultimate goal is to allow mobile phone to correctly recognize all words spoken by user in real time independent of vocabulary size, noise, speaker characteristics or accent.

The Punjabi ASR System is built in the training phase while recognition performance is evaluated during the testing phase. The major portion of the speech corpus is

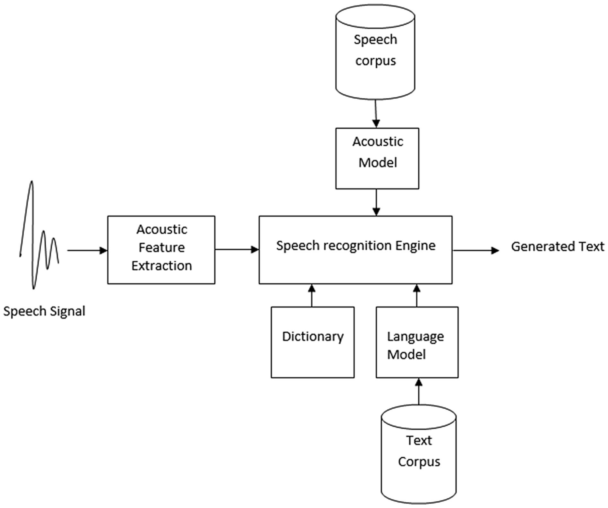


Fig. 1. Components of proposed system.

used to train the system while rest of the recordings is used for testing purpose. Training phase covers speech and text corpus preparation, acoustic feature extraction, dictionary preparation, acoustic model development and language model development. These are finally used by speech recognition engine for text generation. Testing phase covers the evaluation of performance parameters like accuracy, error rate, speed and space utilization for the developed system. This section covers the training phase in detail while the testing phase is discussed in results section.

6.1 Text and Speech Corpus Preparation

Text corpus is the prerequisite for the language model. Text corpus for Punjabi consists of 10 digits (0 to 9) for phone number and two commands 'saaf karo' (to clear number) and 'dial karo' (to dial number). Speech corpus is the prerequisite for acoustic model development. Speech corpora required for acoustic model are not available for Punjabi. The speech corpus for Punjabi is designed to satisfy a set of criteria, which specify the required quality of speech data and the proportional distribution of data with different speaker characteristics. Table 4 provides the technical specifications of the speech recordings.

The speech corpus is representative of native speakers of the Punjabi who are comfortable in speaking and reading the language. The speakers having all the diversities attributing to the gender, age and dialect are chosen for recordings. Every speaker has its own style of speaking, especially male and female voices are quite different. Figures 2 and 3 show waveform representation of digits 0-9 for male and female voices. Male speakers are generally having higher pitch and frequency than the female speakers. Speech recordings of 50 speakers are recorded for 10 digits (0 to 9) and two commands- 'saaf karo' (to clear) and 'dial karo' (to dial). Out of these recordings, the training set consists of 6 h 25 min of speech from 35 speakers (18

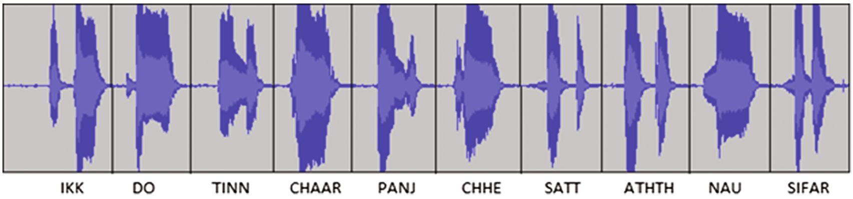


Fig. 2. Waveform of 10 Punjabi digits in male voice.

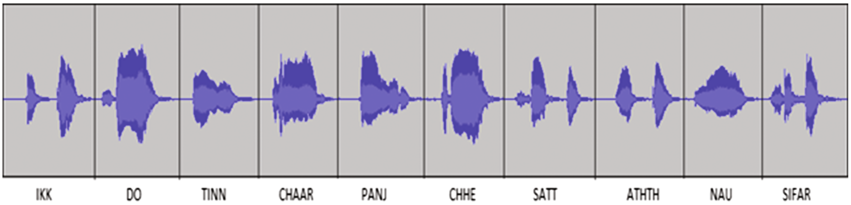


Fig. 3. Waveform of 10 Punjabi digits in female voice.

Table 4. Technical Details of Recordings.

Parameter	Value
Sampling rate	16 kHz
Number of bits	16
Number of channels	1, Mono
Audio data file format	.wav
Corpus	Punjabi 10 digits and 3 words
Number of speakers	50
No. of male speakers	25
No. of female speakers	25
Range of age group of speakers	18–35 years
Average recording time per speaker	11 min per speaker ~9 h for all speakers
Number of recordings per speaker	118
Total number of recordings	5900
Size of raw speech	1010 MB
Condition of noise	Normal life
Window type	Hamming, 25.6 ms
Frames overlap	10 ms

Table 5. Punjabi dataset description.

Data set	Number of tokens	Number of speakers	Speakers' gender		Total number of recordings	Recording time
Punjabi_corpus (training)	13 (10 digits and 3 words)	35	18 female	17 male	4130	6 h 25 min
Punjabi_corpus (testing)	13 (10 digits and 3 words)	15	7 female	8 male	1770	2 h 45 min

female and 17 male) while testing set comprises of 2 h 45 min of speech from 15 speakers (7 female and 8 male). Mobile phone is used to collect recordings having minimal background disturbance. Speakers were asked to utter digits in sequence as well as at random for better accuracy and the recordings were stored in.wav files. Any mistakes made while recording have been undone by re-recording or by making the corresponding changes in the transcription set. Table 5 provides details of training and testing data sets.

6.2 Acoustic Feature Extraction

The training starts with the process of feature extraction. It is one of the most important and crucial steps in speech recognition. In this step parametric and acoustic-phonetic speech features are extracted from the recordings and stored in.mfc file. The unwanted and redundant speech signals are removed to improve the recognition accuracy and pre-processed necessary speech signals are forwarded to the speech recognition engine. The acoustic feature consists of first and second derivatives of 13 dimensional Mel Frequency Cepstral Coefficients (MFCC). The window size of 25 ms and frame shift of 10 ms is considered for MFCC.

6.3 Pronunciation Dictionary

ASR relies on the comprehensiveness of pronunciation dictionary that maps words to their corresponding pronunciation forms in terms of their phonetic representation in a specific language. As discussed earlier, pronunciation dictionary for Punjabi may have word-based, sub-word-based or character-based phonetic representation. So, three pronunciation dictionaries are created for the proposed system.

The word-based dictionary consists of 13 words, Sub-word based dictionary consists of 22 sub-words and character-based dictionary consists of 24 unique characters,

Table 6. Rules Pronunciation of Punjabi digits.

Punjabi Digits in English	Digits in Numerical form	Digits in Punjabi – Word based Phonetic Representation	Character based Phonetic Representation	Sub word based Phonetic Representation
Ikk	1	ਇੱਕ	ਇ ◌ੱ ਕ	ਇੱ ਕ
Do	2	ਦੋ	ਦ ◌ੋ	ਦੋ
Tinn	3	ਤਿੰਨ	ਤ ਿ◌ ◌ੰ ਨ	ਤਿੰ ਨ
Chaar	4	ਚਾਰ	ਚ ◌ਾ ਰ	ਚਾ ਰ
Panj	5	ਪੰਜ	ਪ ◌ੰ ਜ	ਪੰ ਜ
Chhe	6	ਛੇ	ਛ ◌ੇ	ਛੇ
Satt	7	ਸੱਤ	ਸ ◌ੱ ਤ	ਸੱ ਤ
Athth	8	ਅੱਠ	ਅ ◌ੱ ਠ	ਅੱ ਠ
Nau	9	ਨੌ	ਨ ◌ੌ	ਨੌ
Sifar	0	ਸਿਫਰ	ਸ ਿ◌ ਫ ਰ	ਸਿ ਫ ਰ

Table 7. Pronunciation of Punjabi words.

Punjabi Words in English	Words in Punjabi (Gurmukhi Script)	Character based Phonetic Representation	Sub word based Phonetic Representation
Saaf	ਸਾਫ਼	ਸ ◌ਾ ਫ਼	ਸਾ ਫ਼
Karo	ਕਰੋ	ਕ ਰ ◌ੋ	ਕ ਰੋ
Dial	ਡਾਇਲ	ਡ ◌ਾ ਇ ਲ	ਡਾ ਇ ਲ

representing 10 digits and 03 words of Punjabi. Tables 6 and 7 show the phonetic representation of Punjabi digits and words in English form, Numerical form, word-based, sub-word-based and character-based.

6.4 Acoustic Model Development

An acoustic model is a file that contains statistical representations of each of the distinct sounds that makes up a word called as phoneme. It represents the relationship between the recorded speech and the phonemes. From the speech corpus, 70% of recordings by 50 speakers have been used as a statistical base from which the acoustic model has been developed. HMM based acoustic model trainer Sphinxtrain has been used to create statistical representations for each phoneme in Punjabi. The words are represented as sequence of phonemes where each phoneme has its own HMM having a sequence of states. From each speech recording, sequence of feature vectors are extracted and computed. The basic 3-state HMM model is used for each Punjabi phoneme having one state for the transition into the phoneme, one for the middle part and one for the transition out of the phoneme which join models of HMM units together in the ASR engine.

6.5 Language Model

Language model is a probability distribution over sequence of words. It is used for searching the correct word sequence by estimating the likelihood of the word based on previous words. CMU – Cambridge statistical language modeling toolkit (2016) has been used to develop language model for Punjabi. Text corpus of Punjabi having digits and commands is used for language model development.

7 Experimental Results

The proposed ASR system is developed having the ability to convert real-time speech into text and recognize the digits and commands spoken by the user. To evaluate the system performance against desired performance parameters testing is performed. Pocketsphinx [8], speech recognition system for hand held devices is used as decoder. From the speech corpus, 30% of recordings by 50 speakers are used for testing purpose. The experiments included training and testing of three acoustic models with three pronunciation dictionaries on different GMMs. Each model has been evaluated for the following performance parameters:

Word Error Rate (WER): It is defined as the sum of word errors divided by the number of reference words. It takes into account three error types: *substitution* (the reference word is replaced by another word), *insertion* (a word is hypothesized that was not in the reference) and *deletion* (a word in the reference transcription is missed). Word error rate can be calculated as:

$$WER = (S + I + D)/N \tag{1}$$

where S, I and D represent substitution, insertion and deletion errors respectively while N is total number of reference words.

Accuracy (%WAcc): It is defined as the percentage of words correctly recognized by the speech recognition system. It can be calculated as

$$\%WAcc = 100 - \%WER \tag{2}$$

where %WER is the percent word error rate.

Build Time: It is the amount of time taken to build the acoustic model from the training data. It starts with the feature extraction and finishes when acoustic model is fully built.

Decoder Speed: It is the average time taken by the acoustic model to recognize a word. It specifies the CPU time taken by the decoder to recognize speech of one second duration. An average speed of 0.02 xRT (Real time) means that the decoder takes 0.02 s of CPU time to recognize speech of one second duration. The speed of decoder increases with decrease in average time.

Memory Size: It specifies the storage space required to store the fully built acoustic model.

ASR model having minimum WER, build time and memory size with maximum accuracy and decoder speed are desirable for optimal performance. Results of the three models are analyzed and compared to recommend optimal model for development of ASR for Punjabi mobile applications.

7.1 Performance Analysis of Continuous Acoustic Model

The results of Continuous acoustic model for different pronunciation dictionaries are shown in Table 8. It can be observed that the model attains maximum accuracy of 97.5% for character-based dictionary having WER of 2.5. The maximum accuracy for word and sub-word based dictionary is 85.6% and 94.5% having WER of 14.4% and 5.5% respectively.

It is worth noting that the accuracy of continuous model initially increases with increase in GMMs but decreases for very high value of GMMs. This happens due to the

Table 8. Continuous models.

GMM	WER (%age)			Accuracy (% age)			Time (mins)			Decoding speed (xRT)			Size (Mb)		
	W	S	C	W	S	C	W	S	C	W	S	C	W	S	C
4	23.1	5.6	**2.5**	76.9	94.4	**97.5**	27	38	**37**	0.02	0.02	**0.03**	0.45	0.37	**0.35**
8	16.7	**5.5**	2.5	83.3	**94.5**	97.5	20	**49**	34	0.02	**0.02**	0.03	0.75	**0.69**	0.68
16	**14.4**	5.6	3	**85.6**	94.4	97	**29**	68	35	**0.02**	0.04	0.03	**1.33**	1.34	1.51
32	14.9	6.6	3.8	85.1	93.4	96.2	47	109	60	0.03	0.07	0.04	2.49	2.63	2.82
64	17.3	8.5	5	82.7	91.5	95	213	187	695	0.04	0.2	0.06	4.83	5.23	5.44
128	19.2	9.7	6	80.8	90.3	94	387	368	1234	0.07	0.22	0.12	9.5	10.2	10.5
256	20.5	10.3	7	79.5	89.7	93	975	667	1845	0.1	0.24	0.2	14.6	15.6	15.8

W-Word based, S-Sub-word based and C-Character based

presence of own set of senone Gaussians in continuous acoustic models, that increases drastically with the increase in GMMs thereby hampering the efficiency of mixture computation. So, it is not advisable to build continuous models above 16 GMMs. The character based continuous model outperforms others in terms of accuracy, WER, build time and space requirements. Its only limitation is that the speed of decoder is low that can be neglected at the price of high accuracy.

7.2 Performance Analysis of Semi-continuous Acoustic Model

The results of Semi-Continuous acoustic model for different pronunciation dictionaries are shown in Table 9.

Table 9. Semi continuous models.

GMM	WER (%age)			Accuracy (% age)			Time (mins)			Decoding speed (xRT)			Size (Mb)		
	W	S	C	W	S	C	W	S	C	W	S	C	W	S	C
4	60.9	45.7	33.4	39.1	54.3	66.6	11	20	21	0.01	0.01	0.01	0.19	0.07	0.042
8	36.2	17.7	14.8	63.8	82.3	85.2	12	25	31	0.01	0.01	0.01	0.21	0.09	0.64
16	23.6	11.0	8.8	76.4	89.0	91.2	17	36	41	0.01	0.01	0.01	0.25	0.13	0.11
32	18.0	5.8	4.2	82.0	94.2	95.8	30	62	71	0.01	0.01	0.01	0.33	0.22	0.20
64	10.7	5.2	3.9	89.3	94.8	96.1	51	114	123	0.01	0.01	0.01	0.49	0.4	0.38
128	**10.4**	5.0	3.3	**89.6**	95.0	96.7	**237**	226	243	**0.01**	0.01	0.01	**0.83**	0.75	0.74
256	12.9	**4.5**	**3.0**	87.1	**95.5**	**97.0**	803	**506**	**490**	0.01	**0.01**	**0.01**	1.44	**1.45**	**1.44**

7.3 Performance Analysis of PTM Acoustic Model

The results of PTM model for different pronunciation dictionaries are shown in Table 10. It shows that the model attains maximum accuracy of 97.5% for character-based dictionary having WER of 2.5 at 8 GMMs. The maximum accuracy for word and sub-word based dictionary is 87.8% and 94.8% having WER of 12.2% and 5.2% respectively.

Table 10. PTM model.

GMM	WER (%age)			Accuracy (% age)			Time (mins)			Decoding speed (xRT)			Size (Mb)		
	W	S	C	W	S	C	W	S	C	W	S	C	W	S	C
4	26.1	9	4.7	73.9	91	95.3	12	32	46	0.02	0.02	0.02	0.21	0.1	0.078
8	16.1	6	**2.5**	83.9	94	**97.5**	18	48	**54**	0.02	0.02	**0.02**	0.25	0.16	**0.14**
16	14.3	**5.2**	2.7	85.7	**94.8**	97.3	30	**85**	91	0.02	**0.03**	0.02	0.33	**0.27**	0.26
32	13.9	6	3	86.1	94	97	65	177	186	0.02	0.03	0.05	0.49	0.5	0.49
64	**12.2**	6	2.9	**87.8**	94	97.1	**107**	275	511	**0.03**	0.05	0.05	**0.81**	0.95	0.79
128	14.8	8.2	3.1	85.2	91.8	96.9	1207	310	634	0.04	0.08	0.08	1.45	1.9	1.87
256	17.5	9.6	3.9	82.5	90.4	96.1	1475	584	1424	0.08	0.2	0.22	3.4	3.6	3.73

Results indicate that the accuracy of PTM model initially increases with increase in GMMs but decreases slightly for higher value of GMMs. Character based PTM model is having high accuracy, low WER, low build time, low decoding speed and low space requirement than others.

The performance analysis for the three acoustic models clearly indicates that results obtained with character-based pronunciation dictionary are consistent and far better than word-based and sub-word-based pronunciation dictionaries. So, it is recommended to use character-based pronunciation dictionary for Punjabi ASR system. Further, in-depth study of the three acoustic models with character-based pronunciation dictionary outcomes maximum accuracy and minimum WER at only 4 Gaussians resulting low storage requirement and build time with high decoding speed for continuous models. The time required to build the PTM model is more than continuous models but decreased decoding speed overshadows time. Small size of PTM model makes it suitable for memory-limited mobile phone applications. Semi continuous models work well at 256 GMM. In comparison to other two models their performance is not much good but their decoding speed is very low which make it usable for real time environment.

8 Conclusions

Mobile phones have become an integral part of our daily life. Numerous applications are being developed to increase the usability of mobile phones. To develop speaker-independent ASR system for Punjabi mobile applications, different acoustic models with different pronunciation dictionaries at different Gaussians are evaluated in this paper. It can be concluded that character-based dictionary is the best fit for the Punjabi while phonetically-tied acoustic model gives optimal performance for different accuracy and reliability parameters. So, a phonetically-tied model with character-based dictionary can be used for development of speaker-independent ASR system for Punjabi based mobile phone applications.

References

1. Bharali, S.S., Kalita, S.K.: A comparative study of different features for isolated spoken word recognition using HMM with reference to Assamese language. Int. J. Speech Technol. **18**(4), 673–684 (2015)
2. Commissioner for Linguistic Minorities, Ministry of Minority Affairs, Government of India. 50th Report of the Commissioner for Linguistic Minorities in India. http://www.nclm.nic.in/shared/linkimages/NCLM50thReport.pdf. Accessed 14 Jul 2016
3. Das, B., Mandal, S., Mitra, P.: Bengali speech corpus for continuous automatic speech recognition system. In: International Conference on Speech Database and Assessments Proceedings, Taiwan, pp. 51–55 (2011)
4. Davis, K.H., Biddulph, R., Balashek, S.: Automatic recognition of spoken digits. J. Acoust. Soc. America **24**, 637–642 (1952)
5. Dua, M., Aggarwal, R.K., Kadyan, V., Dua, S.: Punjabi automatic speech recognition using HTK. Int. J. Comput. Sci. **9**(4), 359–364 (2012)
6. Ho, T.H., Liu, C.J., Sun, H.: Phonetic State Tied-Mixture tone modeling for large vocabulary continuous mandarin speech recognition. In: Sixth European Conference on Speech Communication and Technology Proceedings, Hungary, pp. 883–886 (1999)
7. Huang, X., Alleva, F., Hon, H.W., Hwang, M.Y., Rosenfeld, R.: The SPHINX-II speech system: an overview. Comput. Speech Lang. **7**(2), 137–148 (1993)
8. Huggins-Daines, D., Kumar, M., Chan, A.: Pocketsphinx: a free, real-time continuous speech recognition system for hand-held devices. In: International Conference on Acoustics, Speech and Signal Processing Proceedings, pp. I-185–I-188. IEEE, Toulouse (2006)
9. Khaira, S.S.: Punjabi Bhasha Viyakarn Ate Bantar (Punjabi). Punjabi University, Patiala (2011)
10. Klatt, D.H.: Review of the ARPA speech understanding project. J. Acoust. Soc. America **62** (6), 1345–1366 (1977)
11. Kumar, K., Aggarwal, R.K.: A Hindi speech recognition system for connected words using HTK. Int. J. Comput. Sys. Eng. **1**(1), 25–32 (2012)
12. Kumar, R.: Comparison of HMM and DTW for Isolated Word Recognition System of Punjabi Language. In: 15th Iberoamerican Congress on Pattern Recognition Proceedings, SP, Brazil, pp. 244–252 (2010)
13. Kumar, Y., Singh, N.: An automatic spontaneous live speech recognition system for Punjabi Language corpus. Int. J. CTA **9**(20), 9575–9595 (2016)
14. Kumar, Y., Singh, N.: An automatic speech recognition system for spontaneous Punjabi speech corpus. Int. J. Speech Technol. **20**(2), 297–303 (2017)
15. Lee, K.F., Hon, H.W., Reddy, R.: An overview of the SPHINX speech recognition system. IEEE Trans. Acoust. Speech Signal Process. **38**(1), 35–45 (1990)
16. Lowerre, B.T.: The Harpy Speech Recognition System. Dissertation, CMU (1976)
17. Mittal, P., Singh, N.: Speech based command and control system for mobile phones: issues and challenges. In: International Conference on Computational intelligence and communication technology Proceedings, pp. 729–732. IEEE, Ghaziabad (2016)
18. Naing, H.M.S., Hlaing, A.M., Pa, W.P.: A Myanmar large vocabulary continuous speech recognition system. In: APSIPA Annual Summit and Conference Proceedings, Hong Kong, pp. 320–327 (2015)
19. Placeway, P., Chen, S., Eskenazi, M.: The 1996 HUB-4 Sphinx-3 system, In: DARPA Speech Recognition Workshop Chantilly Proceedings (1996). http://www.itl.nist.gov/iad/mig/publications/proceedings/darpa97/pdf/placewa1.pdf. Accessed 09 Sept 2016

20. Punjab Population Census data. http://www.census2011.co.in/census/state/punjab.html. Accessed 14 Jul 2016
21. Punjabi Language, Encyclopedia Britannica Online. https://www.britannica.com/topic/Punjabi-language. Accessed 05 Jul 2016
22. Satori, H., ElHaoussi, F.: Investigation Amazigh speech recognition using CMU tools. Int. J. Speech Technol. **17**, 235–243 (2014)
23. Schalkwyk, J., Beeferman, D., Beaufays, F.: Google search by voice: a case study. In: Advances in Speech Recognition: Mobile Environments, Call Centers and Clinics Proceedings, pp. 61–90. Springer (2010)
24. Schuster, M., Nakajima, K.: Japanese and Korean voice search. In: International Conference on Acoustics, Speech, and Signal Processing Proceedings, pp. 5149–5152. IEEE, Kyoto (2012)
25. Thangarajan, R., Natarajan, A.M., Selvam, M.: Syllable modeling in continuous speech recognition for Tamil language. Int. J. Speech Technol. **12**, 47–57 (2009)
26. Walha, R., Drira, F., El-Abed, H., Alimi, A.M.: On developing an automatic speech recognition system for standard Arabic language. Int. J. Electr. Comput. Energ. Electron. Commun. Eng. **6**(10), 1138–1143 (2012)
27. Wang, H.M., Ho, T.H., Yang, R.C.: Complete recognition of continuous Mandarin speech for Chinese language with very large vocabulary using limited training data. IEEE Trans. Speech Audio Process. **5**(2), 195–200 (1997)

ReviseOnTheGo – Immersive Mobile Applications for Accelerated Mathematics Learning

Atharva Kimbahune[1](✉), Sanjay Kimbahune[2],
and Snehal Kimbahune[2]

[1] Ramrao Adik College of Engineering, Navi Mumbai, India
atharva@reviseonthego.com
[2] Thane, India

Abstract. There is a lot of study related stress on school students. Consequences of such stress drive some students to take drastic actions such as suicide [3]. With thorough analysis, we found that Mathematics, is one of the major source of stress for most of these students. To help them relieve their stress to some extent, an immersive suite of mobile applications, named as "ReviseOnTheGo", was developed by the first author [Atharva], who just passed out his SSC exam a few years back. He was supported by his mother, who is trained singer. She lent her voice for the formulas. There are a number of courses online, but they are typically long and one has to complete them in order to benefit. For revision, what students need are precise formulas. This is the main focus of the suite ReviseOnTheGo. Till date there are $\sim$25000 direct downloads from the Google Play Store. This number could be inaccurate because of 3rd party app stores and also due to use of popular app sharing apps. Many students have acknowledged that these simple mobile apps have helped them to score better marks whilst being stress free. This concept also demonstrates, how students can create very useful and relevant mobile applications for themselves as well as other students. All of these downloads were achieved without any major marketing campaigns. ReviseOnTheGo received a special jury mention in the prestigious IAMAI awards (2015) [7] and were runner-ups in the final round of Aegis Graham Bell Awards (2014) [8]. Times Of India newspaper wrote an article on these apps [1]. In this paper, the journey of application development, the challenges faced, learning, and a scale-up plan, are described in detail.

Keywords: Android · ReviseOnTheGo · Mobile app · Educational stress · Formula revision · MIT AppInventor

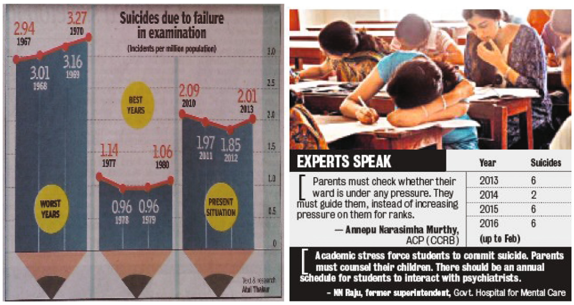
1 Introduction

We have seen the news about students killing themselves due to educational stress, especially after board results.

S. Kimbahune—Social Innovator Thane India
S. Kimbahune—Home maker Thane

S.M. Thampi et al. (eds.), *Advances in Signal Processing and Intelligent Recognition Systems*, Advances in Intelligent Systems and Computing 678,
DOI 10.1007/978-3-319-67934-1_34

India lost 2471 young citizens in 2013 as they committed suicide after failing in their examinations. This has increased by around 10%, when compared to the previous year. Over 2246 students took their lives after they failed examinations in 2012. Maharashtra topped the list with 349 suicides, followed by Tamil Nadu at 277 and Andhra Pradesh being at 235, according to the latest National Crime Record Bureau (NCRB) report [4] (Fig. 1).



EXPERTS SPEAK

Parents must check whether their ward is under any pressure. They must guide them, instead of increasing pressure on them for ranks.

— Annepu Narasimha Murthy, ACP (CCRB)

Year	Suicides
2013	6
2014	2
2015	6
2016 (up to Feb)	6

Academic stress force students to commit suicide. Parents must counsel their children. There should be an annual schedule for students to interact with psychiatrists.

– NN Raju, former superintendent, Govt. Hospital for Mental Care

Fig. 1. Reference from Times of India and Indian express

We thought of doing something helpful for these unhappy and stressed friends, so that they do not hurt themselves. First author had passed out the SSC exam in 2012 and had a firsthand experience of this stress. Thus, being aware of what exactly the students would want, created a range of Android applications for revising formulas and theorem on the go. Between various subjects, their difficulty, it was concluded that the most common source of stress was subject of “Mathematics.” The problem is that students forget the formulas at the time of exam. This happens purely due to the lack of revision. Thus, the concept of “ReviseOnTheGo” was thought of. These applications are essentially mobile based audio-visual revision applications, which were rapidly developed using MIT AppInventor Framework.

2 Vision

Vision of ReviseOnTheGo program is to reduce the educational stress of students studying up to graduation. Also, through these suite of applications, reach at least 1 million students by December 17. This is envisioned as a social movement with a snowballing effect on students and to enable them to create such applications.

3 Problem Definition

Looking at various subjects, their difficulty level and psychological pre-conceived notions; it was concluded that the most common source of stress was subject “Mathematics.” Mathematics means formulae, theorem and their proofs. Most of the time Murphy haunts! One does not remember the exact formulae when needed.

The main problem is; students forget the formulas at the time of exam. This is purely due to lack of revision. Thus, the concept of "ReviseOnThe Go" was thought of.

4 The Solution: Revision App

The most common solution; writing the formulas on a piece of paper. It surely helps, but isn't very effective. Moreover, it is not convenient to carry a bunch of sheets everywhere. Another approach is to click photographs of formulas on the phone and revise whenever time permits. But this still not convenient and effective, as one has to zoom into it every time, has quality issues depending on the camera of the phone and so on. So, it was thought to use using clear text. But it was not so much different than the previous thing. Thus, arrived at conclusion of making Audio-Visual applications [1, 5]. As of now there are 3 applications. Algebra app displays and speaks 30 formulas. There are two apps of Revise Geometry. One is for formulas of theorems and other being the proofs. First app of geometry displays and speaks 80 formulas and the other part displays 27 theorems and their proofs.

It is as easy as listening to music. When one listens to music repeatedly, one tends to remember the lyrics of the song quickly [1]. These applications are used best while travelling or when one is at peace. The user has to use ear-phones/head-phones and listen to the formulas as many times as wanted, and bingo – the formulas get internalized.

5 Development Experience and Challenges

Various alternatives for android application development were evaluated. Coding was one of the alternative however it needed special training. Looked at rapid application development tools and shortlisted MIT's AppInventor. AppInventor, is a cloud-based drag and drop android application development environment. One needs to select the appropriate blocks, connect them to other blocks and configure them with respect to the desired functionality. These blocks are compiled and an .APK file (android package) is generated.

A design document that depicted call flow and text was created first and was reviewed by few students pursuing SSC at that time. This document was used as reference to configure the call flow. Tested the generated APK file on few devices and were also distributed to few close friends. During this process, many issues were discovered and were fixed. The issues were related to typos, choice of image format not working on certain mobiles, etc. Once all the issues were resolved, it was hosted on Google Play Store.

Some of the challenges faced while developing the app are listed below.

1. Managing application size: Since this app is audio visual, the images and audio files initially took considerable amount of space. Initially, the images which were large in size and weren't supported by many mobile phones. This issue was discovered after launch. Later on, this bug was fixed in the AppInventor and thus the issue was

also fixed. Audio files were recorded in .wav format. Though it was uncompressed audio, it made the applications' size too large and inconvenient. So, after converting them to.amr format, the size was brought under control.

2. Designing an UI that is appealing: The first author being a student and had just passed out SSC, put in lot of effort to make UI relevant to the theme of these applications.
3. Mistakes while configuring the flow: By mistake one parameter was set out of limits (max volume of the audio file). The app used to get compiled, but then, when run on the phone, it crashed. Contacted the support group forum, they tried to help but could not figure out the bug. Then used android debug bridge (ADB) and figured out that the app was maxing out the CPU on the android's main thread. The app was just trying to set the conflicting value and could not succeed. This happened repeatedly. So, the main thread was overloaded, and android force closed it every time. After, referring to AppInventor's guide, finally found that the volume has a range of 0 to 100 only, while it was set it to 150 with the intention of increasing the audibility. Finally, the property value was changed, and the problem was solved. Thus, the learning from this mistake, is to be careful while deciding the parameters to be set. These small things aren't generally documented well. Such simple mistake took a lot of efforts and energy to fix.

6 Business Model

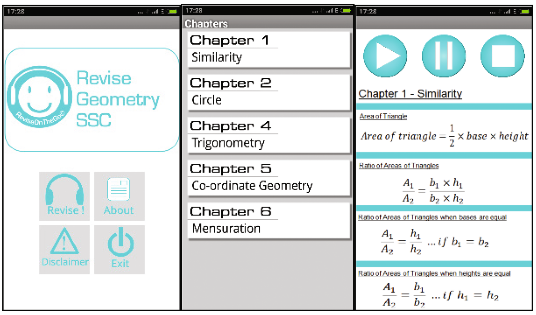
These apps were never intended to make more money; however, it has to be self-sustaining. Understanding the fact that it would be difficult to charge students, the following alternatives are thought of: 1. Integrate advertisements in the application. 2. Look for funding from CSR's of corporate companies. 3. Look for angle VC supporting social businesses. 4. Create an easy method for voluntary donations.

7 Scale up Model

Leverage Apex bodies of science teachers in India. Convince them and let them be the front runners. Talk to publishers like Navneet so they can put an advertisement in their guides or provide them with a white labeled application for customization. Connect with NGO'S who are working in field of popularizing Mathematics and Sciences, especially with students living in rural areas. Leverage social media to reach more students. Create a framework where in, it is easy for students to develop applications for themselves and others. This could be video of how to make apps using simple mobile app development tools. Also, looking at 3rd party app distribution channels (like AppsDaily, etc.). Suggest use of off-line distribution methods and attempt to make it as a part of the preinstalled apps in modern smartphones.

8 Screenshots of the Apps

See Figs. 2, 3, 4 5 and 6.



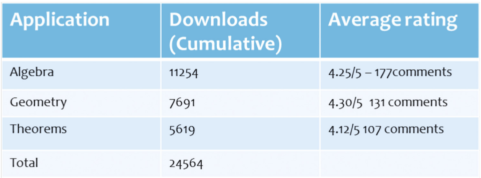
$Area\ of\ triangle = \frac{1}{2} \times base \times height$

$\frac{A_1}{A_2} = \frac{b_1 \times h_1}{b_2 \times h_2}$

$\frac{A_1}{A_2} = \frac{h_1}{h_2} \ldots if\ b_1 = b_2$

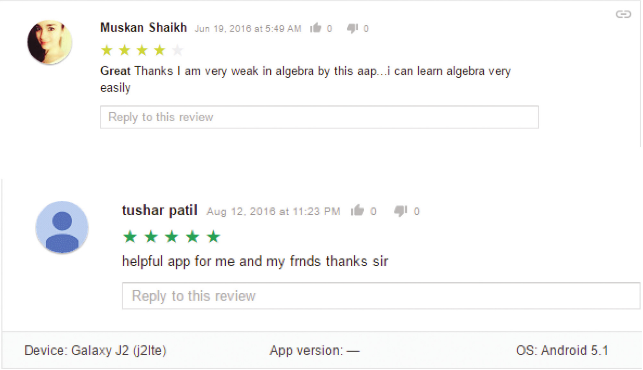
$\frac{A_1}{A_2} = \frac{b_1}{b_2} \ldots if\ h_1 = h_2$

Fig. 2. Screenshots for Geometry App.



Application	Downloads (Cumulative)	Average rating
Algebra	11254	4.25/5 – 177comments
Geometry	7691	4.30/5 131 comments
Theorems	5619	4.12/5 107 comments
Total	24564	

Fig. 3. Google Play Store downloads.



Muskan Shaikh Jun 19, 2016 at 5:49 AM 0 0

★★★★☆

Great Thanks I am very weak in algebra by this aap...i can learn algebra very easily

tushar patil Aug 12, 2016 at 11:23 PM 0 0

★★★★★

helpful app for me and my frnds thanks sir

Device: Galaxy J2 (j2lte) App version: — OS: Android 5.1

Fig. 4. Feedback from students

http://tinyurl.com/ssc-algebra

http://tinyurl.com/ssc-geometry

http://tinyurl.com/ssc-theorem

Fig. 5. Resources and links for downloading the apps

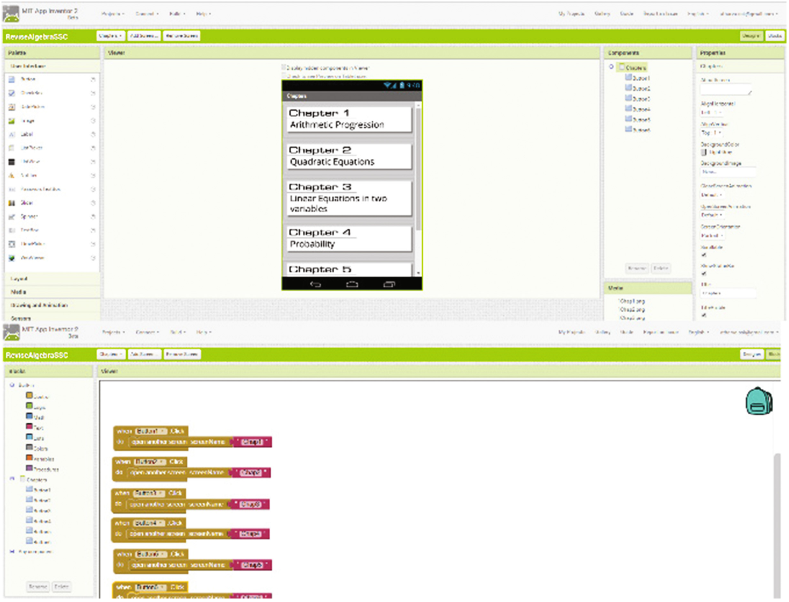


Fig. 6. Representative Screenshots of application development environment of MIT AppInventor

9 Conclusion

Students can make compelling and good applications. Frameworks such as “MIT AppInventor” makes it very easy to develop mobile apps. Student even need not know coding at all. These applications (3 nos.) were developed with less than 3 months efforts by a SSC student. His mother, who is a trained singer and a coauthor, lent her voice for vocalizing the formulas.

For revision, one needs the precise formulas itself. Vocalization of formulas coupled with on screen display is the key to quickly revise and internalize the formulas. Tone and the voice modulation of the recordings play an important role in the adaptation by a student. The tone of recording was kept as if mother is teaching a student rather than a teacher teaching student. This was one of the major factor for scaled adoption.

The new generation is extremely attached to their phones and to music. This concept leverages the strengths of these two, to positively assist and influence the younger generation to study. Headsets enable students to revise the formulae wherever they go or whatever the loud noisy condition they may be in. Typically, at the time of festive seasons, there are disturbing loud noises and the students have a hard time focusing. With these apps and with a headset, the student gets isolated from background noise and can focus on revising. This is true even when the student is travelling in a bus or a train

Headsets enable students to revise the formulae wherever they go or whatever the loud noisy condition they may be in. Typically, at the time of festive seasons, there are disturbing loud noises and the students have a hard time focusing. With these apps and with a headset, the student gets isolated from background noise and can focus on revising. This is true even when the student is travelling in a bus or a train.

Acknowledgements. Author would like to thank following people who significantly contributed and encouraged this initiative. This includes Mr. Surendra Dighe of Jidnyasa Trust, IAMAI, Bell Awards, Mr. Prasad Patki and all the students who used the apps and posted encouraging comments.1

References

1. http://timesofindia.indiatimes.com/city/thane/18-yr-old-comes-up-with-app-for-last-minute-SSC-revision/articleshow/51664433.cms
2. http://www.reviseonthego.com
3. http://www.newindianexpress.com/states/andhra_pradesh/Stress-Drives-Students-to-Commit-Suicide/2016/03/20/article3336686.ece
4. http://www.dnaindia.com/mumbai/report-maharashtra-tops-student-suicide-list-2002506
5. http://appinventor.mit.edu/explore/
6. Mobile multimodal applications on mass-market devices: experiences Regensburg, Germany, 3–7 September 2007. University of Washington (2007). ISBN 0-7695-2932-1Number: UMI Order No. GAX95-09398
7. https://www.youtube.com/watch?v=NojYPeMifWk
8. https://www.youtube.com/watch?v=tCWedltMrg4 (fast forward to 31:24)

Analysis of the Electric Arc Furnace Workshop Logistic Processes Using Multiagent Simulation

Konstantin Aksyonov, Anna Antonova(✉), and Natalia Goncharova

Ural Federal University, Mira, 19, Ekaterinburg, Russia
wiper99@mail.ru, antonovaannas@gmail.com,
n.v.goncharova@urfu.ru

Abstract. This paper describes analysis of the electric arc furnance workshop logistic processes using knowledge-based multiagent simulation. The goal of the investigation is to decrease the average time of melts processing by changing melts grouping parameters before supplying to the shop input. A multiagent simulation model has been created in order to analyze various variants of the melts grouping. Agents have a knowledge base that contains an information about logic of the of the melts movement in various routes. The following decision has been revealed. The melts grouping with identical routes before supplying to the shop input allows decreasing an average queue time and average time of melts processing.

1 Introduction

In metallurgical production, the efficiency of the workshop is determined not only by the volume of production produced per unit time, but also by the following factors: the average duration of processing unit of production in the workshop depending on the type of route, the average loading of the workshop aggregates, the average queue of products before processing on aggregates, the number of units of production assigned to another order per unit of time. When analyze the data of the electric arc furnace (EAF) workshop, it is convenient to analyze the above-mentioned characteristics of the shop's operation using a simulation model of the workshop. The use of simulation technology allows to significantly reduce costs at the stage of various reconstructions of the workshop, as well as to determine the production capacity of the equipment under various operating conditions of the workshop [4, 5].

In the various production logistic problem studies, the various advantages of the simulation method are revealed. Dynamic simulation allows the generation of many logistical operations in production for a long time. As a result, the operation statistics are collected and bottlenecks are identified. The bottlenecks of the metallurgical manufacture connected with the cranes and ladle cars movement has been investigated using simulation in [5]. Simulation allows evaluating the alternative system operation by changing the input model parameters in different experiments. Alternative variants of the reconstruction of the metallurgical workshop have been evaluated from the point of view of the cost and volume of output with the help of simulation in [4]. The logistic task of the vehicles distribution along the existing routes, taking into account the

S.M. Thampi et al. (eds.), *Advances in Signal Processing and Intelligent Recognition Systems*, Advances in Intelligent Systems and Computing 678,
DOI 10.1007/978-3-319-67934-1_35

vehicles capacity and the costs of their movement, has been solved using simulation in [7]. This experience can be generalized in the case of the vehicles (cranes and ladle cars) movement in the shops of the metallurgical enterprise.

Development of the knowledge-based multiagent simulation model of EAF workshop logistic processes affecting the workshop efficiency is relevant. We consider the development of the EAF workshop model with the help of the module for creating models of a metallurgical enterprise information system. The metallurgical enterprise information system is a knowledge-based web-oriented decision support system for tracking, monitoring, modeling, analysis, and improvement processes of the steel products manufacturing [1, 3].

The remainder of the paper is organised as follows: Sect. 2 formulates the electric arc furnance workshop logistic processes problem. Section 3 describes a multiagent model of EAF workshop and suggests recomendations. Section 4 evaluates the experiment results. Section 5 concludes paper and explores future work.

2 Problem Formulation

We consider the EAF workshop consisting of the following aggregates: one electric arc furnace aggregate, one steel finishing aggregate (SFA), one ladle-furnace aggregate (LFA), and one continuous casting machine (CCM). The processing on the EAF aggrsegate proceeds consistently and continuously accords to the sequence numbers of the melts. In total, 24 operations are performed on the EAF aggregate, and 6 types of unplanned operations are possible.

The melt in the EAF workshop passes 4 different routes depending on the purpose of the final unit of production. Conditionally, we denote these routes by letters.

Route 'A' EAF–CCM. The percentage of the melts with this route is 25.3% of the total number of the melts. The time for moving of the melt between the EAF aggregate and CCM is 4 min. Further, the melt is processed on CCM according to it serial number (waits until all the previous melts have been processed).

Route 'B' EAF–LFA–CCM. The percentage of the melts with this route is 9.7% of the total number of the melts. The time for moving of the melt between the EAF aggregate and LFA is 1 min. Further, the melt on the average 4 min awaits treatment at LFA. Then, in the order of the queue (according to the sequence numbers of the melts), the melt is processed at LFA. The time for moving of the melt between LFA and CCM is 2 min. Further, the melt is processed on CCM according to it serial number.

Route 'C' EAF–SFA–CCM. The percentage of the melts with this route is 38.8% of the total number of the melts. The time for moving of the melt between the EAF aggregate and SFA is 2 min. Further, the melt on the average 5 min awaits treatment at SFA. Then, in the order of the queue (according to the sequence numbers of the melts), the melt is processed at SFA. The time for moving of the melt between SFA and CCM is 2 min. Further, the melt is processed on CCM according to it serial number.

Route 'D' EAF–LFA–SFA–CCM. The percentage of the melts with this route is 26.2% of the total number of the melts. The time for moving of the melt between the EAF aggregate and LFA is 1 min. Then, the melt on the average 4 min awaits treatment at LFA. Then, in the order of the queue (according to the sequence numbers of the melts), the melt is processed at LFA. The time for moving of the melt between LFA and SFA is 1 min. Further, the melt on the average 5 min awaits treatment at SFA. The time for moving of the melt between SFA and CCM is 2 min. Further, the melt is processed on CCM according to it serial number.

Processing at LFA. The processing proceeds according to the processing queue, which is formed from the sequence numbers of the melts. 1% of the melts is re-processed on LFA, and such melts after the primary treatment on LFA are placed at the end of the general melts queue for processing on LFA. 4.5% of the melts after processing on LFA is assigned to another order (reassigned).

Processing at SFA. The processing proceeds according to the processing queue, which is formed from the sequence numbers of the melts. 1% of the melts is re-processed on SFA, and such melts after the primary treatment on SFA are placed at the end of the general melts queue for processing on SFA.

Processing at CCM. The casting proceeds according to the casting queue, which is formed from the sequence numbers of the melts. CCM works continuously. Since the melts on CCM is processed sequentially according to their ordinal number, the current melt is waiting for processing on CCM until all the previous melts have been casted.

The problem for optimization of logistic processes is formulated as follows: it is necessary to simulate the operation of the EAF workshop for 10 days in a 24-hour mode. It is necessary to find the best values of the output characteristics of the shop's operation depending on the following input conditions.

1. Melts that submitted for processing on EAF are mixed regardless of the route (basic experiment).
2. Before feeding to EAF, the melts are grouped into 20 melts with the identical route. The sequence of routes is the following: A–B–C–D.
3. Before feeding to EAF, the melts are grouped into 20 melts with the identical route. The sequence of routes is the following: D–C–B–A.
4. Before feeding to EAF, the melts are grouped into 20 melts with the identical route. The sequence of routes is the following: C–A–B–D.
5. Before feeding to EAF, the melts are grouped into 40 melts with the identical route. The sequence of routes is the following: A–B–C–D.
6. Before feeding to EAF, the melts are grouped into 40 melts with the identical route. The sequence of routes is the following: D–C–B–A.
7. Before feeding to EAF, the melts are grouped into 40 melts with the identical route. The sequence of routes is the following: C–A–B–D.

3 Development of the Simulation Model of EAF Workshop

The EAF workshop model has been developed using a notation of multiagent resource conversion processes [2]. Agents in the model are used to implement the logic to process the orders for production, determine the orders route, adjust the experiment and create the probability of events. Model operations describe the duration of the operation of the EAF workshop elements in order to assess the queues and to find the experimental parameters. In the model developed, one order is described. Order z1 'Melt' passes through the model according to the plan for melt processing. After passing the set working cycle, the order is deleted. Figure 1 shows the structure of decomposition of the EAF node.

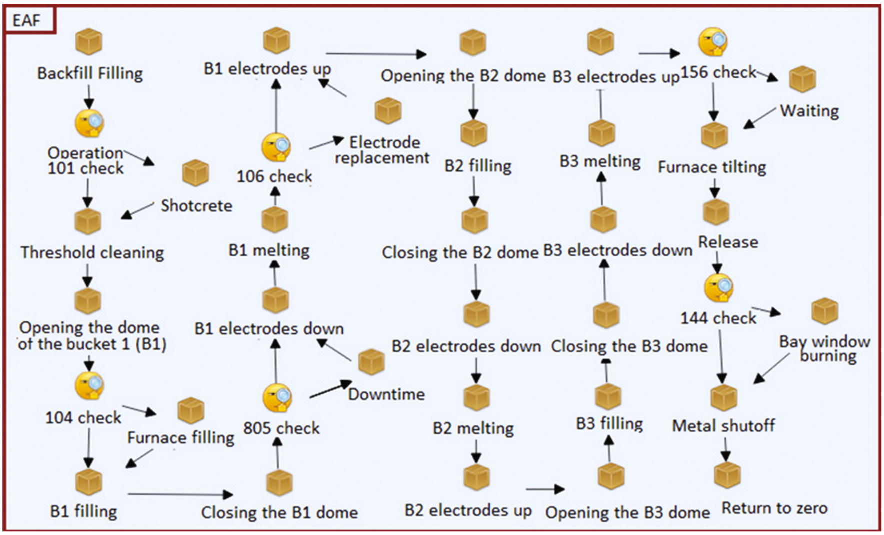
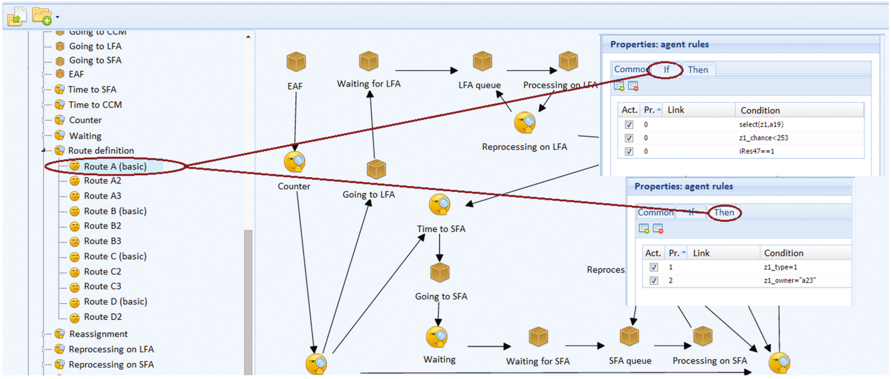


Fig. 1. Decomposition of the EAF node

Agents are used to determine the occurrence of unplanned operations depending on their occurrence probability.

Distribution of the melts along different routes is carried out with the help of 'Route definition' agent. The distribution algorithm is described using knowledge-based model of the form 'If-Then' (Fig. 2). The knowledge-based modeling is used to describe the experience of the decision makers on the different levels of the workshop work.



Properties: agent rules

Common | If | Then

Act.	Pr.	Link	Condition
☑	0		select(z1,a19)
☑	0		z1_chance<253
☑	0		iRes47==1

Properties: agent rules

Common | If | Then

Act.	Pr.	Link	Condition
☑	1		z1_type=1
☑	2		z1_owner="a23"

Fig. 2. Example of description of the knowledge base rule of 'Route definition' agent

The model structure can be divided into four work units: (1) description of the EAF operation including description of the route selection process for melt; (2) description of the LFA operations; (3) description of the SFA operations; (4) description of the CCM operations.

In this model, 24 operations are performed during the melt processing on EAF and 6 types of unplanned operations are possible. The description of these operations in the model is implemented in the decomposition of the EAF node.

For the experiments, the ‘Experiments’ agent is used, which depending on the experiment number, changes the necessary settings (the number of the melts in the group and the sequence number of the route).

The developed model of the EAF workshop in comparison with the existing model [4] is expanded by 24 operations of an electric arc furnace, among which there are unplanned downtime operations related to the various violations in the technological process of steelmaking.

4 Analysis of Experiments Results

We consider the experiments with the developed model of the EAF workshop in the optimization module of the metallurgical enterprise information system. The simulation time discrete was selected equal 1 min per tact; therefore, to simulate the work of the workshop for 10 days, the simulation experiment was conducted during 14400 tacts.

Upon completion of the basic experiment, a detailed report in the format.csv was created containing detailed information on changing the values of all model parameters with the course of model time. With the help of this report, the graphics of the basic experiment have been built (Fig. 3). In figure, the x-axis is modeling time in minutes and the y-axis is average queue in melts, waiting processing on aggregate per time unit.

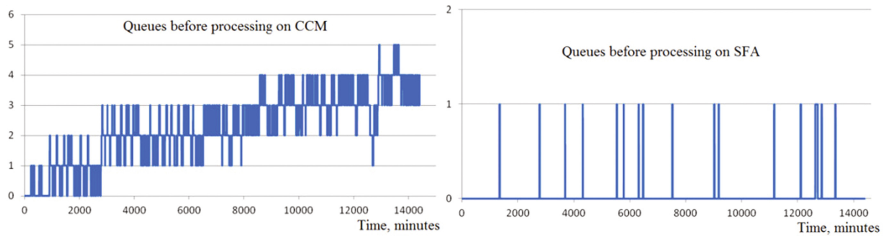
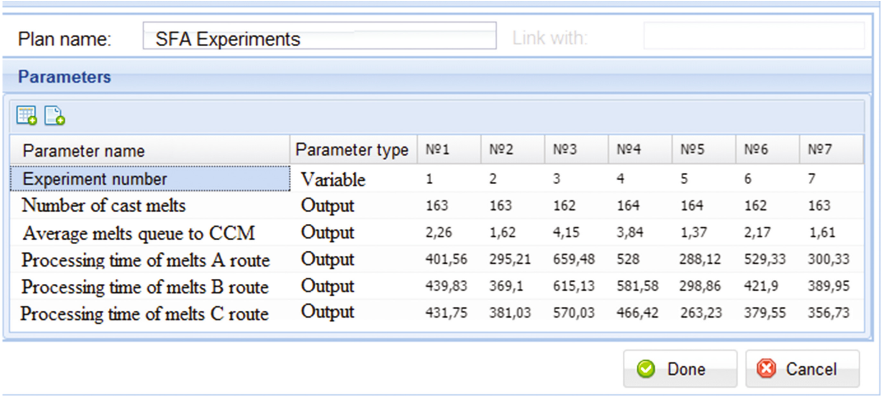


Fig. 3. Simulation results for the basic experiment: queues of the melts

As follows from the analysis of Fig. 3, the most significant queues of the melts occur before processing on CCM; the queues before processing on SFA and LFA do not exceed one melt.

In order to determine the most efficient way of operating the EAF workshop, it is necessary to determine the order of delivery of melts, in which the number of cast melts will be the largest and the values of the workshop output characteristics will be optimal. For this purpose, a plan of experiments was set up and executed in the optimization module of the metallurgical enterprise information system (Fig. 4). As the workshop output characteristics, the following were chosen: the number of cast melts, the average waiting times for melt processing on CCM and SFA in minutes, the average value of the melts queue ahead of CCM, the average processing time of the one melt depending on the route type in minutes.



Plan name: SFA Experiments

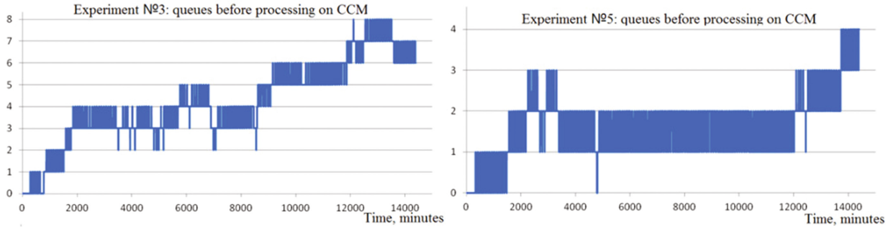
Link with:

Parameters

Parameter name	Parameter type	№1	№2	№3	№4	№5	№6	№7
Experiment number	Variable	1	2	3	4	5	6	7
Number of cast melts	Output	163	163	162	164	164	162	163
Average melts queue to CCM	Output	2,26	1,62	4,15	3,84	1,37	2,17	1,61
Processing time of melts A route	Output	401,56	295,21	659,48	528	288,12	529,33	300,33
Processing time of melts B route	Output	439,83	369,1	615,13	581,58	298,86	421,9	389,95
Processing time of melts C route	Output	431,75	381,03	570,03	466,42	263,23	379,55	356,73

Fig. 4. View of the executed plan of experiments in the optimization module

The experiment with the best result is experiment number 5 because in this experiment the queue of the melts waiting processing on CCM is the lowest. In this experiment, the number of cast melts is one of the largest (164 melts); in addition, for this experiment there is a decrease in the average processing time of the one melt by 18–26 % depending on the route type. Reducing the average processing time of the one melt is caused by a decrease for the experiment №5 the average waiting times for melt processing on CCM by reducing the melts queue ahead of CCM (Fig. 5). The average waiting time for melt processing on CCM for experiment №5 does not exceed 15 min.



Experiment №3: queues before processing on CCM

Experiment №5: queues before processing on CCM

Fig. 5. Simulation results for experiments № 3 and № 5 : queues of the melts

Thus, the following conclusion has been obtained. The best overall performance of the EAF workshop is achieved by grouping the melts before serving on EAF with 40 melts with the identical route and using the following sequence of routes: EAF–CCM, EAF–LFA–CCM, EAF–SFA–CCM, and EAF–LFA–SFA–CCM.

To solve the logistic problems of production in similar studies [4] and [5], modeling systems AnyLogic [6] and Plant Simulation [7] have been applied respectively. In these studies, an assessment of alternative modes of the melts movement along technological routes has been made using simulation experiments. The best mode of the melts movement has been revealed among analysing modes. The advantage of the AnyLogic system is using agents to describe the system elements behaviour. The advantage of the Plant Simulation is formation of detailed template reports on the statistics of the system operating. The advantage of the simulation module of the metallurgical enterprise information system is integration of the multiagent simulation and system of the collection and storage of production data obtained from production sensors and used in the simulation.

5 Conclusion and Future Work

In this paper, the knowledge-based multiagent simulation model of the EAF workshop logistic processes has been described in the module for creating models of the metallurgical enterprise information system.

The developed model of the EAF workshop operation has been applied to solve the problem of optimizing the supply of melts for treatment in the workshop depending on the melt route type and the number of grouped melts with the identical route. The following conclusions have been obtained. When using the melts grouping at the EAF workshop input, the average awaiting time before CCM is reduced and, accordingly, the average time for melt treatment in the EAF workshop is reduced; the rest output characteristics of the workshop are changed slightly. Based on the results of the analysis of experiments with the simulation model of EAF workshop, the following recommendations have been proposed. It is necessary to group the melts before EAF workshop for 40 melts in the group with the identical route using the following sequence of routes: EAF–CCM, EAF–LFA–CCM, EAF–SFA–CCM, and EAF–LFA–SFA–CCM.

Further development of simulation models for metallurgical production with the help of the metallurgical enterprise information system is promising from the point of view of analyzing and optimizing the logistic processes of the enterprise.

Acknowledgments. The work was supported by Act 211 Government of the Russian Federation, contract № 02.A03.21.0006.

References

1. Aksyonov, K., Bykov, E., Aksyonova, O., Antonova, A.: Development of real-time simulation models: integration with enterprise information systems. In: Proceedings of the Ninth International Multi-conference on Computing in the Global Information Technology, pp. 45–50 (2014)
2. Aksyonov, K., et al.: Decision support systems application to business processes at enterprises in Russia, efficient decision support systems – practice and challenges in multidisciplinary domains. In: Jao, C. (ed.) InTech, pp. 83–108 (2011). http://www.intechopen.com/articles/show/title/decision-support-systems-application-to-business-processes-at-enterprises-in-russia, ISBN 978-953-307-441-2
3. Borodin, A., Kiselev, Y., Mirvoda, S., Porshnev, S.: On design of domain-specific query language for the metallurgical industry. In: Proceedings of 11th International Conference BDAS: Beyond Databases, Architectures and Structures: Communications in Computer and Information Science, pp. 505–515 (2015)
4. Chelyabinsk metallurgical plant uses a simulation model of electric-furnace melting shop. http://anylogic.com/case-studies/chelyabinsk-metallurgical-plant-uses-a-simulation-model-electric-furnace-melting-shop
5. Klebanov, B., Mufazalov, A., Myasoedov, I., Krymov, E.: Use of plant simulation for improvement technological and business processes of metallurgical manufacture. In: Proceedings of the 35th Chinese Control Conference, pp. 9681–9684 (2016)
6. Mutiagent modeling system AnyLogic. The official web site. http://www.anylogic.com
7. Osaba, E., Carballedo, R., Diaz, F.: Simulation tool based on a memetic algorithm to solve a real instance of a dynamic TSP. In: Proceedings of the IASTED International Conference Applied Simulation and Modelling, pp. 27–33 (2012)
8. Plant Simulation official web site. https://www.plm.automation.siemens.com/en/products/tecnomatix/manufacturing-simulation/material-flow/plant-simulation.shtml

Development of Intelligent Petrol Supplies Planning System

K.A. Aksyonov[1(✉)], A.L. Nevolina[1], H.L. Ayvazyan[1,2],
and O.P. Aksyonova[1]

[1] Ural Federal University n. a. the First President of Russia B.N. Yeltsin,
Ekaterinburg 620002, Russia
bpsim.dss@gmail.com, hambardzum.ayvazyan@gmail.com
[2] Yerevan State University, 0025 Yerevan, Armenia

Abstract. This work describes results of development of intelligent petrol supplies planning system. There have been made an analysis of transportation problem. For solve transportation problem of petrol station network was suggested to integrate some methods and approach (transportation algorithm, expert system, multi-agent approach and simulation modeling) for taking into account main features of petrol supplies planning and scheduling problem. Frame-based approach has been used for subject area formalization. Supposed method of decision making for supplies scheduling problem and intelligent system are currently implicating on petrol stations network in Ekaterinburg city and Sverdlovsk region.

Keywords: Modeling · Simulation · Multi-agent · Transportation problem · Intelligent system · Petrol station · Networks

1 Introduction

One of spheres of computer modeling is the transport logistics - the section of logistics which is directly connected with the organization of movement of material flows. In modern market conditions the transport logistics plays an important role as any enterprise is interested in the most optimal movement of objects for economy of resources.

One of the applied directions of Multi agent technologies is planning [3]. The concept of an agent corresponds to hardware or software implemented entity, which is capable of acting for the benefit of goal achievement set by the owner or user. Agents possess certain intelligent capabilities [3, 7]. The Multi agent system for logistics is provided in [4].

A sample application of the multi-agent system for planning operation of a flexible production system is discussed in [7]. We may name the following advantages of the multi-agent system:

(1) Formalization of decision making points in form of the agents. The points include specific situation processing scenarios. Technically this process is a part of knowledge formalization stage.

S.M. Thampi et al. (eds.), *Advances in Signal Processing and Intelligent Recognition Systems*, Advances in Intelligent Systems and Computing 678,
DOI 10.1007/978-3-319-67934-1_36

(2) Planner is dynamically embedded by means of interaction of specific element of the multi-agent system and thus is ready to modify the plan in case of delays or unexpected (unintended) situations. The planner works in real-time.
(3) Agent network, interconnected with relations, self coordinates its activity.

An additional benefit of multi-agent planning is the capability of automated information sharing between process individuals about changes of controlled object, which introduces control transparency. Subject area knowledge is being formalized during development and deployment of the planning multi-agent system, the decision making process is automated. Thus we ease activity, related to decision making. In this work simulation is used for bottleneck analysis [8, 15].

For the purpose of collection of basic data the analysis of planning processes of transportations and a delivery of fuel between petroleum storage depots (OB), stationary and automatic gas stations (AGS and AAGS) Ekaterinburg and its satellites in Sverdlovsk region of the fuel Ergo brand was carried out. This fuel network consists of 38 AGS and AAGS, and AGS of acquirers which are supplied with fuel from several petroleum storage depots. The auto fleet uses as own gasoline tank trucks, and uses services of freelancers (hired gasoline tank trucks) participating in a delivery of fuel.

The model consisting of the interconnected two contours was developed for debugging of a method and algorithms of planning: the multi-agent simulation model (developed in system of dynamic modeling of situations BPsim. MAS) and the intellectual agent of the planning realized in system of decision support BPsim. DSS [5, 6]. At the heart of the BPsim complexes the model of multi-agent process of transformation of resources is programmatically realized. In this work the task of search of the effective plan of a delivery of fuel is solved on the integration of classical transport task and multi-agent simulation.

2 Application of a Transport Task

The task of planning of a delivery of fuel comes down to a classical transport task. A transport task (Monge-Cantorovitch’s task) - a mathematical task of linear programming of a special type about search of optimal distribution of uniform objects from the accumulator to receivers with minimization of costs for movement [1].

There are suppliers and consumers of some uniform freight. Each supplier has a certain quantity of units of this freight (capacity of the supplier). Each consumer needs a quantity of units of this freight (demand of the consumer). Carriage costs of an unit of cargo from each of suppliers to each of consumers are known. The main goal of a transport task – to constitute such plan of transportations in case of which total costs on cargo hauling will be minimum [2]

The transport system of the company on providing oil products contains the following components:

(1) gas stations (gas station, automatic gas stations (AAZS));
(2) park of fuel trucks (truck and hired fuel trucks);
(3) oil depots (NB, and/or others which contracts on providing fuel and lubricants are signed);

(4) the list of routes of fuel trucks movement from oil depots to gas station;
(5) the list of the types of fuel realized by the company.

The companies on providing oil products can have on hand the oil depots or use services of others. The cost of purchase of oil products on third-party oil depots is higher, than own oil depot of the company, but involvement of third-party oil depots is required for the following reasons:

– a number of gas station of the company is at considerable remote distance from the oil depots so fuel on more relatives to data of gas station fuel warehouses is more favorable to buy the companies;
– some types of fuel realized at gas station of the company are absent on the oil depots therefore the companies are required additional sources of fuel for providing network;
– owing to complexity of process of supply of oil depots (interruptions in deliveries from oil refineries (oil refinery) and unpredictability of the schedule of transportation on the railroad) and also mismatches in dynamics of consumption of fuel on network of gas station and volumes of deliveries to the oil depots from oil refinery, it is vital to interact to small and average networks of gas station with oil depots of larger players (most often it is the vertically integrated oil companies (VINKi)).

The park of fuel trucks of the company may contain its own vehicles or for the solution of the tasks to attract hired fuel trucks (freelancers). Fuel trucks have various capacities of a tanker truck and consumption rate of fuel. Each fuel truck, depending on brand, has several sections as a part of a tanker truck and can transport from 1 to 6 types of various fuel (the quantity of sections at small and medium-sized companies varies from 1 to 3, however large networks and VINKi also use 6 section fuel trucks).

During the research the main algorithms of the solution of the closed task model on examples of transportation of fuel were considered such as the method of a northwest corner, the method of a minimal cost and method of distribution. The following situation was taken as an example: suppliers (sections of gasoline tank trucks) A1, A2, A3 and A4 concentrated respectively 5960, 4440, 6010 and 4420 L of some fuel which is necessary to deliver to consumers (to reservoir at the relevant gas stations and AAGS) B1, B2, B3 in desirable quantity 5000, 10500, 4330 L. Classical parameter in a transport task "cost" is determined through a priority of transportation and set by a matrix:

$$\mathrm{P} = \begin{pmatrix} 2,1 & 2 & 2 \\ 1 & 2,5 & 1 \\ 1 & 2,8 & 2 \\ 2,8 & 2,1 & 1 \end{pmatrix}$$

where the priority of transportation is the number of days which the gas station will be able to work uninterruptedly without new supplies for each type of fuel. The less is the priority– the more is the request for a supply urgent.

The method of a northwest corner is expected introduction of delivery to the upper left corner of the table until of the consumer demand is satisfied or supplier capacity

Table 1. The result of the method of the north-west corner

Section volume, liter	The required volume, liter		
	6000	10500	4330
	Transportation priority/Transportation volume, liter		
5960	2,1/5960	2/	2/
4440	1/40	2,5/4400	1/
6010	1/	2,8/6010	2/
4420	2,8/	2,1/90	1/4330

isn't exhausted. Otherwise, this consumer or the supplier is ignored. Such plan can't be optimal as the freight charges aren't considered. The result of work of an algorithm is provided in Table 1.

The method of a minimal cost is in that on each step delivery with the smallest priority (cost) of transportation of an unit of cargo among all blank cages was chosen. This method allows to find the plan near optimal in case of small time of search. The result of work of an algorithm is provided in Table 2.

Table 2. The result of the method of minimal cost

Section volume, liter	The required volume, liter		
	6000	10500	4330
	Transportation priority/Transportation volume, liter		
5960	2,1/	2/5960	2/
4440	1/	2,5/110	1/4330
6010	1/6000	2,8/10	2/
4420	2,8/	2,1/4420	1/

The distribution method assumes checking whether the plan received by other method is optimal by means of creation of a matrix of estimates. If this plan isn't optimal, the optimization cycle then the following checking on an optimality is made.

This method also carries the name of a method of potentials and is the most exact, but at the same time, the most labor-consuming. The most exact plan of cargo hauling for the reviewed example is provided in Table 3. This result is received by means of four cycles of optimization.

$$\text{Matrix of estimates: } O = \begin{pmatrix} 1,9 & 0 & 1,8 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \\ 2,5 & 0 & 0,7 \end{pmatrix}.$$

Table 3. The result of the distribution method

Section volume, liter	The required volume, liter			Estimate
	6000	10500	4330	
	Transportation priority/Transportation volume, liter			
5960	2,1/	2/5960	2/	−0,2
4440	1/	2,5/110	1/4330	−1
6010	1/6000	2,8/10	2/	−1
4420	2,8/	2,1/4420	1/	−0,3
Estimate	0	−1,8	0	

Job analysis of algorithms of the solution of a transport task allows to draw a conclusion that the algorithm that most suitable for further development is the method of a minimal cost. This approach is taken in a basis of an algorithm of planning of a delivery of fuel for intellectual system.

To the additional factors limiting application of a transport task to a task of a delivery of fuel is the following:

(1) the frequency rate of amount of cargo hauling shall be multiple to section amount;
(2) the freights aren't uniform and each freight (depending on a fuel type) can be transported in one section of the gasoline tank truck;
(3) the sequence of discharge of fuel by the gasoline tank truck (depending on design features of drain devices the priority of discharge of sections can differ) isn't considered;
(4) there is no time component in the form of times of the beginning and the end of flights, loading/unloading times;
(5) there is no separation into types of freights or their marking (fuel types for example: 92, 95, 98, Dt, 80);
(6) the presence at the gasoline tank truck of several sections isn't considered;
(7) physical restraints of gasoline tank trucks on servicing of gas station aren't considered;
(8) there is no opportunity to consider preferences of gasoline tank trucks on servicing of gas station;
(9) the possibility of servicing of close gas stations by one gasoline tank truck for flight isn't considered.

In logistic field multi agent approach is used in Magenta technology. Magenta technology based on the nets of requirements and capabilities [13, 14]. But clearly muti agent approach guaranty only rational decision but not optimal.

Results of the analysis of approaches to the solution of a problem of planning for oil products supply of network of gas station are generalized in Table 4.

Table 4. Analysis of approaches to the solution of problems of deliveries planning

Criteria\methods	Transport task	Imitating modeling	Expert systems
Questions of adequacy of model			
Use of means (motor transport)	+	+	NO
Streams of resources (volumes of transportations): NB-AGS, NB-Truck-AGS	+/NO	+/+	+/+
Time of transportation, loading, plan of delivery	NO	+	NO
Decision-maker model heuristics, agents of planning	NO	NO	+
Support of the solution of tasks			
Planning taking into account restrictions - times/resources/means	NO/+/+	+/+/+	NO/+/+
Analysis of bottlenecks of processes	NO	+	NO
Dispatching	NO	+	NO

Multi agent systems are used for problems of planning and management in real time, is focused on search of rational decisions for problems of big dimension. As the transport task is not focused on the solution of problems of planning flights and drawing up routes and also does not consider a number of requirements, in this work the transport task is used at a stage of distribution of volume of deliveries from oil depots to gas station (without binding to fuel trucks). Other factors are offered to be considered with use of Multi agent approach. In the following section is carried out the analysis of two Multi agent approaches: networks of requirement and opportunities (the PV-networks considered in Sect. 1.5) and dynamic modeling of Multi agent process of resources transformation.

3 Multi-agent Simulation in BPsim Suite

For the description of a simulation model 3 types of requests are entered into subject domain: the gasoline tank truck, the supply requisition, general settings (for a task of parameters of an experiment).

The Petroleum storage depots block imitates activities for gas station of gasoline tank trucks and supply of petroleum storage depots. Agents of the block “All gas stations” imitate “consumption” of various type of fuel from everyone, described in model, reservoir. The “consumption” changes depending on time of day: decreasing at night (from 23 h to 7 h) and increasing in day (from 7 h to 23 h). The Management of Requests block imitates work of the flight controller (logistics specialist) performing the analysis of a condition of reservoirs of gas station. The current level of reservoirs of all gas stations, and, depending on the strategy of a delivery is analyzed, creation of supply requisitions of fuel is performed if the current level of fuel is already not enough for consumption on the set number of days.

Directly the planning algorithm, is realized in the intellectual agent in the BPsim. DSS system using frame model of representation of knowledge in the basis.

Knowledge representation model based on frame concept and conceptual graphs of Sowa J. F. [9–12]. The intellectual agent has the visual interface and two operating modes: automatic and automated. In case of situations, the flight controller can reconstruct completely the plan of a delivery or make corrections to it.

Further management is transferred to the agent "Control of gasoline tank trucks". It "checks" availability of the free gasoline tank truck, and in the presence of requests, the gasoline tank truck "is reserved" for delivery of products for target AGS. In the presence of several requests for one AGS, and in the presence of several sections at the gasoline tank truck, several sections for gasoline supplies for several reservoirs at one AGS are reserved directly, and several supply requisitions are processed.

Decision support system is implemented on the basis of BPsim.DSS software suite [5, 6, 8]. The comparison of experiment result of new method and real plan is presented on Table 5.

Table 5. Comparison analysis of experiment result

Date	Number of haul		Volume of delivery	
	Fact	Experiment	Fact	Experiment
22.09	24	22	330955	303530
23.09	7	14	105504	249972
24.09	22	20	303036	304103
25.09	16	18	242312	267848
28.09	21	21	338067	320061
29.09	23	19	344424	289223
30.09	20	19	251136	289761
02.10	18	29	264748	440392
03.10	18	23	275179	349396
04.10	20	20	303533	305164
Summary	189	205	2758894	3119450

Results of computing experiments of a simulation model are compared with actual data of a delivery and showed convergence of results regarding flights and amount of transportation of fuel.

4 Conclusion

As a result of a research the simulation model and decision support system of gas stations network has been developed, on the basis of technology of program agents the algorithm of creation of the plan of a delivery is realized, experiments are made. Results of computing experiments of imitating model are compared with actual data of a delivery and have shown convergence of results regarding flights and the volume of transportation of fuel.

Acknowledgements. The work was supported by Act 211 Government of the Russian Federation, contract № 02.A03.21.0006.

References

1. Galyutdinov, R.R.: Transportation problem – decision of potential method. http://galyautdinov.ru/post/transportnaya-zadacha
2. Prosvetov, G.I.: Mathematical Method in Logistic: Problem and Decision, p. 304. Alfa Press, Moscow (2014)
3. Wooldridge, M.: Agent-based software engineering. IEEE Proc. Soft. Eng. **144**(1), 26–37 (1997)
4. Kowalski, M., Zelewski, S., Bergenrodt, D., Klupfel, H.: Application of new techniques of artificial intelligence in logistics: an ontology-driven case-based reasoning approach. In: Proceedings of ESM 2012 (ESM - European Simulation and Modelling Conference) 22–24 October 2012, FOM University of Applied Sciences, pp. 323–328. Essen, Germany (2012)
5. Aksyonov K., Bykov E., Aksyonova O., Goncharova N., Nevolina A. Decision Support for Gasoline Tanker Logistics with BPsim.DSS. International Conference on Computer Information Systems and Industrial Applications (CISIA 2015). June 28-29, Bangkok, Thailand. pp. 604–606. WOS:000359866200164
6. Aksyonov, K., Bykov, E., Aksyonova, O.: Petrol delivery management with BPsim.DSS. In: 33rd Chinese Control Conference, CCC 2014, Nanjing, China; 28–30 July 2014, pp. 7628–7632 (2014)
7. Jennings, N.R.: On agent-based software engineering. Artif. Intell. **117**, 277–296 (2000). http://www.agentfactory.com/~rem/day4/Papers/AOSEJennings.pdf. Accessed Jun 2014
8. Aksyonov, K., Bykov, E., Aksyonova, O., Goncharova, N., Nevolina, A.: Extension of the multi-agent resource conversion processes model: implementation of agent coalitions. In: 5th International Conference on Advances in Computing, Communications and Informatics, ICACCI 2016, Jaipur, India, pp. 593–597 (2016). doi:10.1109/ICACCI.2016.7732110, http://www.scopus.com WOS:000392503100094
9. Sowa, J.F.: Conceptual graphs for a database interface. IBM J. Res. Dev. **20**(4), 336–357 (1976)
10. Sowa, J.F.: Conceptual Structures: Information Processing in Mind and Machine, p. 481. Addison-Wesley, Reading (1984)
11. Sowa, J.F.: Knowledge Representation: Logical, Philosophical, and Computational Foundations, p. 594. Brooks/Cole Publishing Co., Pacific Grove (2000)
12. Aksyonov, K., Bykov, E., Smoliy, E., Sufrygina, E., Aksyonova, O., Kai, W.: Development and application of decision support system BPsim.DSS. In: Proceedings of the IEEE 2010 Chinese Control and Decision Conference (CCDC 2010), 26–28 May 2010, Xuzhou, China, pp. 1207–1212 (2010)
13. Wittich, V.A., Skobelev, P.O.: Multi-agent interaction models for the design of the nets of requirements and capabilities in open systems‖. Aut. Telemechanics **1**, 177–185 (2003)
14. Skobelev, P.O.: Multi-agent technologies for the control of resources in real-time. In: Proceedings of the Seminar —Mechanics, Control and Informatics‖, Track —Perspective Computer Systems: Devices, Methods and Concepts, Tarusa, 2-4 March 2011. http://www.iki.rssi.ru/seminar/2011030204/presentation/20110303_03.pdf. Accessed 13 Apr 2016
15. Devyatkov, V.V., Vlasov, S.A., Devyatkov, T.V.: Cloud technology in simulation studies: GPSS cloud project. In: Proceedings of the 7th IFAC Conference on Manufacturing Modeling, Management, and Control, IFAC, Saint Petersburg (Russia), 19–21, June 2013, part 1, vol. 7, pp. 637–641 (2013)

A Model for an Emotional Respondent Robot

Ayşe E. Sancar[1(✉)] and Elena Battini Sönmez[2]

[1] Department of Psychology and Department of Computer Engineering,
Istanbul Bilgi University, Istanbul, Turkey
ece.sancar@bilgiedu.net
[2] Department of Computer Engineering,
Istanbul Bilgi University, Istanbul, Turkey
elena.sonmez@bilgi.edu.tr

Abstract. The aim of this study is to design an emotional regulation model based on facial expressions. It is argued that emotions serve a critical function in intelligent behavior and some researchers posed the questions of whether a robot could be intelligent without emotions. As a result, emotion recognition and adequate reaction are essential requirements for enhancing the quality of human robot interaction. This study proposes a computational model of emotion capable of clustering the perceived facial expression, and using cognitive reappraisal to switch its internal state so as to give a human-like reaction over the time. That is, the agent learns the person's facial expression by using Self Organizing Map, and gives it a meaning by mapping the perceived expression into its internal state diagram. As a result, the presented model implements empathy with the aim to enhance human-robot communication.

1 Introduction

Intelligent agent start having a close touch with all walks of life with important applications in several fields such as gaming, advertisement, health and personal assistance.

Focusing on the health field, researches support the idea of a robotic assistant for children with autism by longitudinal research [1, 2]; also, service robots could be beneficial in retirement houses to give various health care for elderly people [3]. Since this kind of interactions require good communication, there is the need to increase emotional and cognitive abilities of agents. We conjecture that empathy is one of the key issues affecting the quality of interaction, and we focus on the study of automatic facial expressions, which gives important cue for assessing the (human) partner's emotional state, and on the implementation of a communication model, which enable the robot to associate facial expressions to its internal states (embodiments of emotions). Having strengthened those connections, the robot will be able to understand the partner's emotion, to synchronize its internal state to the one of the human partner, and to move to the most suitable next emotional state. In other words, the aim of this study is to model a coherent course of emotions over the time.

Several architectures have been proposed for artificial emotions and some of the models are equipped with animated face and body. While some models focus on

S.M. Thampi et al. (eds.), *Advances in Signal Processing and Intelligent Recognition Systems*, Advances in Intelligent Systems and Computing 678,
DOI 10.1007/978-3-319-67934-1_37

mechanical engineering with aim to mimic the input facial expression back to the spectator, others try to implement a model to control the emotional states of the humanoid robots.

In 2003, Breazeal [4] presented the emotional model of the animated face KISMET. The top-level goal of the agent is to satisfy its three drives, the social, the stimulation and the fatigue ones, by bringing them to their homeostatic regimes. The drives and the external stimulus sensed by KISMET affect its current set of beliefs, which are then mapped into a 3D (arousal, valence, stance) affective space to produce an emotion. The trigged emotion is finally displayed by the motor and the behavior systems.

In 2003, Arkin et al. [5] introduced the AIBO dog and the SDR humanoid robots. Both agents are equipped with a 3D mental model of emotions having activation, pleasantness and certainty dimensions and an ethological model which provides a basis for behavior selection.

In 2004, J. Gratch and S. Marcella [7] presented their computational model of emotion, EMA, based on the appraisal theory. EMA keeps an explicit representation of the agent-environment relationship, called ‘causal interpretation’, which consists of a snapshot of the agent’s current knowledge. External stimulus as well as agent’s behavior change this knowledge over the time. During the ‘appraisal derivation’ stage all significant features of the causal interpretation are represented into a data structure, or ‘appraisal frame’, which are passed to the ‘emotion derivation’ model to produce an emotional response, the agent’s coping strategy. The triggered emotional response is biased by an overall mood state.

In 2004, Miwa et al. [8] developed the WE-4R human-like robot capable of communicate naturally with human by expressing human-like emotions.

In 2007, Esau et al. [6] introduced the robot head MEXI, which uses artificial emotions and drives to control its behavior. MEXI considers the four basic emotions of anger, happiness, sadness and fear.

In 2007 Hirth et al. [9] presented a behavior based emotional control architectures for the robot head ROMAN.

In 2007, Watanabe et al. [10] proposed an agent which learns to link the expression of the human partner with its own internal state via intuitive parenting.

In 2009, Hasimoto et al. [11] developed a head robot called KAMIN to enhance human-robot communication. The analysis of the human voice allows KAMIN to recognize the human emotion, which is then mapped into the Russell’s circumflex 2D model. The emotional generation space of KAMIN realizes the entrainment between the human and the robotic emotion.

In 2014, Angelica Lim and Hiroshi G. Okuno [12] presented their computational model of emotion based on the field of developmental robotics [13]. The main assumption is that emotions are not built-in, but they can be learned by any learning entity (a child or a robot), when it receives the correct stimuli. The emotion developmental system designed in this paper is implemented into the Multimodal Emotional Intelligence (MEI) robot.

Although several computational models of emotions have been proposed before, there is still the need to investigate on this issue and to create alternative architectures, which can also be used to investigate psychological theories.

This paper proposes a new emotional respondent robot, which can regulate itself regarding the flow of the dialog with a human partner, by reading user's facial expression and using cognitive reappraisal to switch to its next artificial internal state; at the end of each episode the agent acknowledges the human's current emotional state and its affective state. The main contribution of this work is to increase the quality of human-robot interaction by adding empathy to our computational model of emotion.

Section 2 gives an overview of the current theories of emotions; Sect. 3 introduces the proposed system with implementation details. Section 4, presents our experimental setup and results. Lastly, interpretation of the results and future works are in Sect. 5.

2 Background for Theory of Emotion

Minimally, a psychological theory of emotion explains the cognitive and social emotional sphere of human being. The presence of several psychological theories of emotions complicates further its implementation, but it adds also value to it, since a computational model can be used to validate the corresponding theory.

Currently, the main three theories of emotions are:

Discrete Theories

They assert that there is a limited number of core emotions and their expressions is shared across people and culture [14]; these emotions are happiness, sadness, surprise, fear, anger and disgust. Each basic emotion serves to prepare and motivate the reaction to a particular context. Moreover, emotions allow to learn new behavior, and they are also refined throughout emotional development. The main criticism to these theory is that it fails to describe the complexity of the emotional space.

Dimensional Theories

Emotions are mapped into 2D and 3D emotional space. Computational models of 3D based on these theories often use the PAD emotional space of Mehrabian and Russell [15], where the three dimensions corresponds to Pleasure (amount of pleasantness i.e. liking versus disliking), Arousal (amount of mental alertness and physical activation, it describes how excited or apathetic is the emotion i.e. sleepiness of boredom versus exciting), and Dominance (a measure of power or control versus submissiveness i.e. anger, boldness, and relaxation versus fear, anxiety and loneliness).

Appraisal Theories

These theories assert that emotions reflect the personal-environment relationships; that is, emotions arise from the process of comparing individual internal needs to the external demands. The result of this evaluation is mapped into a set of appraisal variables that produce an emotional response [16]. Furthermore, in 1991, Lazarus [17] details how the trigged emotion leads to coping responses, which can be "problem-directed", when they aim to change the environment, or "emotion-directed", when they aim to modulate the trigged emotion. This process of emotion stabilization is called behavioral homeostasis [18]. The major criticisms to these theories argue that emotions are mainly reactive, and appraisal can be considered as a consequent, not a precedent of the emotional reaction. Moreover, humans experience 'feelings', and mood, as well as personality affect the trigged emotion. That is, appraisal rules are insufficient to explain all the complexity of the emotional reaction.

3 Design of the Proposed Model

3.1 Overview

The virtual agent introduces in this paper monitors the user by looking at him-her, and recognizing his-her mimics, as depicted in Fig. 1. Moreover, it gives meaning to the user's expression via mapping it into its internal state diagram; which is designed regarding the three dimensions of pleasure, arousal and dominance axes [19]. After the observation, the agent regulates its affect state dynamically considering the user's facial signals and its current state. In positive and negative circumstances, the agent uses cognitive reappraisal to keep away itself from a high level emotional moods and keep itself in more positive emotion. Cognitive reappraisal allows the agent to change its current state considering its internal and desired states. At the end of the regulation process, the agent gives a statement about its new emotional state. That is, the output of one regulation phase depends on three matters:

1. Which affect state the agent is in
2. What is the message of the user's expression
3. The values of α and β used in Eq. (1)

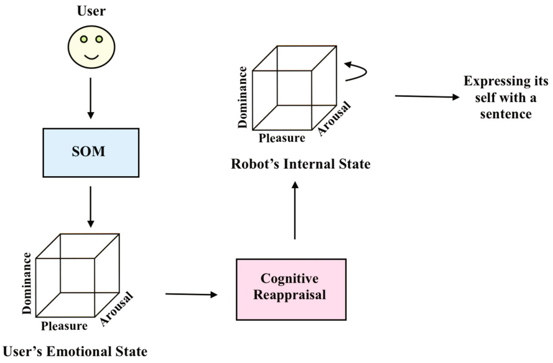


Fig. 1. Regulation steps of the agent's internal state in a dialog with a user regarding facial expressions

3.2 Cognitive Reappraisal

People use emotional regulation strategies consciously or nonconsciously, when they have too high or too low emotional state. A purpose of it is to make changes on their emotional response. One of these strategies is cognitive reappraisal and it brings about changing on focused other potential aspect of the circumstances. In other words, reappraisal shrinks the possible emotional effect of circumstance.

For instance, a student who did not study enough for her/his exam blames her/his teacher to ask too hard questions [20, 21].

There are more four different emotional regulation strategies which are situation selection, situation modification, attentional deployment and response suppression. Each strategy focuses on a different element to change. Through the studies, cognitive reappraisal and response suppression were found two most effective strategies. Furthermore, using cognitive reappraisal has more positive effect on socialization, since it does not cause to hide negative or positive social clues and also it affects both emotional reaction and experience [21]. Therefore, in this paper cognitive reappraisal is considered.

In this model, the agent adjusts the effect of internal and external perception by decreasing its impact. After regulated its effect, new values of the perceived emotional state are considered as desired state. Cognitive reappraisal affects in different level positive and negative situations. If the current recognized state is a negative emotion, intensity of the emotion decreases at rate of one quarter of it. On the other hand, when the current state is a positive emotion, this percentage is ten percent of the emotion. By this way, the desired emotional situation is obtained as more positive than the perceived negative situation. The virtual agent's response is determined regarding its current internal state and the next desired state. In this way, the agent reacts suitable for also its current internal state. Equation (1) shows how the virtual agent responses with cognitive reappraisal:

$$internal\_s_{t+1} = \beta * internal\_s_t + \alpha * desired\_s_t \tag{1}$$

Where both β and α are constant values; they determine the ratio of the components of the new emotional state. For instance, if α is 1.0 and β is 0, the agent generates its next affect state in the direction of the emotion which comes from the user.

3.3 Internal State Diagram

The internal state of the agents is designed regarding to computational models of emotion. Most convenient theory is Russell and Mehrabian three-dimensional affect model to represent an internal state. The three-dimensional model is an extension of an two dimension one, where the dominance dimension is added to arousal and pleasure axes (PAD), see Fig. 2 [22], since only two axes are not enough to discriminate all affects. For example, even anger and fear have completely different behavior tendencies and facial expressions, both are on the top left corner on the two-dimensional psychological judgement space which are high arousal and displeasure [23].

In our study, emotions which have dominance value greater than zero were combined on the same surface of internal state, while emotions with dominance value smaller than 0 were mapped into another surface of internal state.

Our synthetic internal state model contains three levels of fear, anger, happiness, sadness emotions and a natural state. Anger and happiness are on the first level of the diagram dominance greater than zero and fear and sadness are on the second level dominance smaller than zero. Besides, both surface involves natural state. Levels of diagram are designed as indicated in Fig. 3:

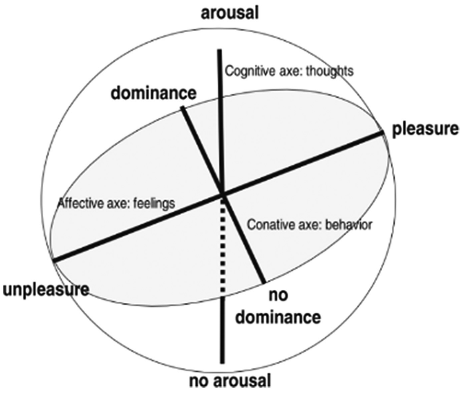


Fig. 2. Diagram of theory of PAD [15]

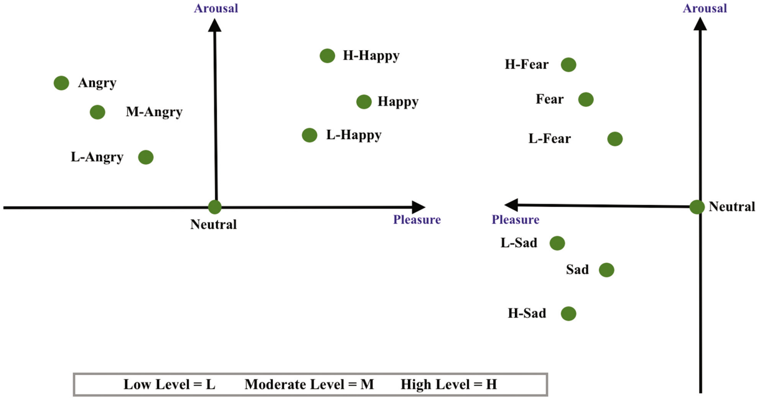


Fig. 3. Internal space of the agent in two surface

To fix the position of every emotions, studies of Mehrabian and Russell [19] and Plutchik [25] are considered. Plutchik had study on semantic and intensity analysis between wide range of emotions; thanks to this research, positions of levels of the four emotions are identified depending on other similar emotions. For example, Plutchik's theory shows that hostility and annoyance are in the similar direction with anger but hostility is lighter than anger and same relationship is valid for annoyance and hostility, annoyance have lower intensity than hostility. In this paper, positions of hostility and annoyance which are obtained from studies of Mehrabian and Russel uses as moderate and low levels of the anger.

3.4 Facial Expression Recognition

The Self Organizing Map (SOM) is used to teach facial expression recognition and the Extended Cohn Kanade Database (CK+) is utilized as training data. The database contains anger (45 faces), contempt (18 faces), disgust (59 faces), fear (24 faces), happiness (69 faces), sadness (28 faces) and surprise (83 faces) emotions. Participants of the database are 210 adults which are ages between 18 and 50 and %67 participants are female. Distribution of participants' ethnic background is %81 Euro-American, % 13 Afro-American and %6 other groups. Participants got training to perform by an experimenter. Each presentation of emotion starts and ends with a neutral pose. From each presentation, picked facial expression is used on the training phase [26].

At first step, four basic emotions of angry, happiness, sadness and fear which are used in the internal states diagram of Fig. 3 and natural poses are chosen and converted to gray scaled images. After face components are detected and cropped as shown in the Fig. 4, images normalized by histogram equalization.

At the second step, vectorized images are given to the SOM to build a map to read the users' facial expressions. We controlled the construction of SOM with the aim to build a 3D clustering space resembling the internal state diagram of Fig. 5: that is, just before the iterative training phase, images which represent the mean for specific category of emotions are assigned at the center positions of each emotion cluster. By doing like that, the resulting SOM (Fig. 7) has a 1-to-1 mapping with the 3D PAD space of Fig. 3 and this gives an opportunity to interpret the emotional state of the user regarding to the internal state diagram of Fig. 3 with the position of the winning node of the emotion reading. That is, the size of the constructed SOM is 8 × 8 × 2 corresponding to 128 emotional states.

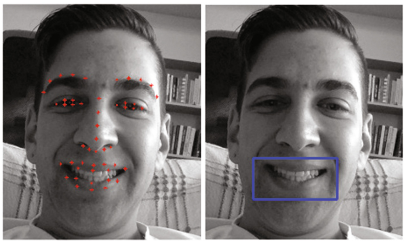


Fig. 4. From left to right: A test image with its face landmarks, the cropped mouth

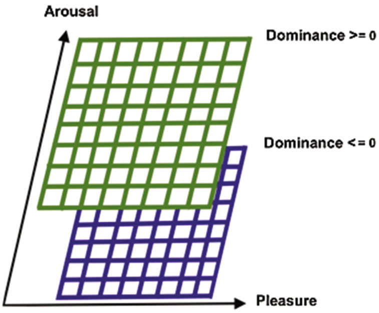


Fig. 5. A representation of the three dimensional emotion clustering

4 Experiment

4.1 Experimental Setup

In the experiment, four states of emotion (happiness, anger, fear and sadness) and natural state are used with three different subject (see Fig. 6). All test images are taken by a camera of the computer in natural environments. Images are presented to the agent in a random way.

At each episode, the agent witnesses the user's facial expression via an image. After the recognition, the user's state of the affect is read on the pleasure – arousal – dominance axes and the rates of dimensions are evaluated regarding the internal state diagram. At the cognitive reappraisal phase, intensity of the current emotional state decreases in specific ratio regarding being negative emotion or not. When re-generating the internal state, seventy percent of the current state of the virtual agent is taken as basis, so the desired situation has an impact on thirty percent of the next state. At the end of each episode, the agent tells how it sees the user and how it feels. Episodes are done consecutively, thus the agent uses the regulated internal state of the previous episode as initial state. Only in the first episode, the agent's initial state is assigned as random inside the pool of natural state.

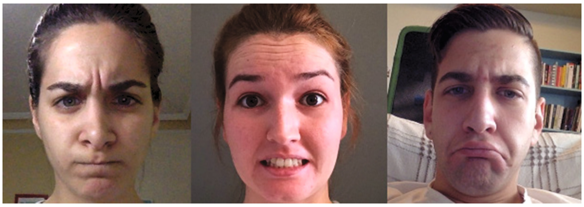


Fig. 6. Example of the users' images. From left to right: subject 1 acting 'angry', subject 2 acting 'fear', and subject 3 acting 'sad'

4.2 Results of Emotion Recognition

According to the PAD rate of the emotional states [10], the ideal clustering map is designed for the five categories of facial expression and they are clustered successfully respecting the ideal map, as shown in Fig. 7. Rows of the maps is representing the pleasure axis, and columns of the maps is representing the arousal axis, the dominance dimension originates the two planes. The central point of the pleasure and arousal axis is between fourth and fifth elements of the 3D SOM. Besides, at each training session, the positions of the clusters do not change due to the control on the map. We worked with the block of the mouth (which is cropped as shown in Fig. 4) since it gave better results than whole face and the block of the eyes. The recognition is done with nearest neighbor.

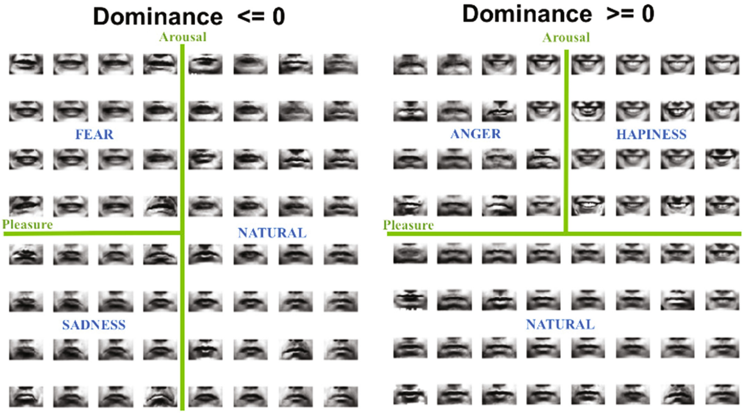


Fig. 7. Output of the self organizing map

4.3 Results of Emotion Regulation

Some examples of the possible episodes are below: the agent acknowledges the current emotional state of the human partner and moves to a suitable internal state. Looking at the first three episodes, we can see that the agent successfully acknowledges the perceived emotional state of the human partner, and empathizes with him-her by switching to a nearby emotional state. In episode four, we can see that the agent struggled to adapt itself to big changes in mood of conversation. That is, the agent gets confused on changing its emotional condition from medium angry to happy. However, except for large transitions, the effect of cognitive reappraisal can be observed more clearly on the rest of the dialog.

Episode 1:
As I see, you feel little sad.
I feel little sad.
Episode 2:
As I see, you feel neutral.
I feel neutral but close to happy.
Episode 3:
As I see, you feel angry.
I feel medium angry.
Episode 4:
As I see, you feel very happy.
I feel angry.

5 Conclusion

In this paper, a design of a respondent emotional robot is presented, also cognitive reappraisal and three-dimensional emotion theory are considered and discussed. The internal state diagram is designed to interpret a user's facial expression and to change the robot's internal state dynamically regarding the flow of conversation. The results are showed that three-dimensional theory is successful to represent the agent's internal state and understand emotional state of a person. Generally, the agent responds properly regarding its current state and external circumstance, except when high volume alterations on the external condition occurs. We believe that this result is consistent to human beings, since people can live more or less the same struggle to jump from a negative situation to a positive one. Future work includes further experiments and discussions as well as quantitative and qualitative evaluation of the respondent robot.

References

1. Kozima, H., Nakagawa, C., Yasuda, Y.: Children-robot interaction: a pilot study in autism therapy. In: Progress in Brain Research, pp. 385–400 (2007)
2. Robins, B., Dautenhahn, K., Boekhorst, R.T., et al.: Robotic assistants in therapy and education of children with autism: can a small humanoid robot help encourage social interaction skills? Univ. Access Inf. Soc. **4**(2), 105–120 (2005)
3. Jayawardena, C., Kuo, I.H., Unger, U., et al.: Deployment of a service robot to help older people. In: IEEE/RSJ 2010 International Conference on Intelligent Robots and Systems, pp. 5990–5995 (2010)
4. Breazeal, C.: Emotion and sociable humanoid robots. Int. J. Hum. Comput. Stud. **59**, 119–155 (2003). doi:10.1016/S1071-5819(03)00018-1
5. Arkin, R., Fujita, M., Takagi, T., et al.: An ethological and emotional basis for human robot interaction. Robot. Auton. Syst. **42**, 191–201 (2003)
6. Esau, N., Kleinjohann, L., Kleinjohann, B.: Integration of emotional reactions on human facial expressions into the robot head MEXI. In: IEEE International Conference on Intelligent Robots and Systems, pp. 534–541 (2007)
7. Marsella, S., Gratch, J.: A domain-independent framework for modeling emotion. J. Cogn. Syst. Res. **5**(4), 296–306 (2003)
8. Miwa, H., Itoh, K., Matsumoto, M., et al.: effective emotional expressions with emotion expression humanoid robot WE-4RII. In: IEEE/RSJ International Conference on Intelligent Robot and Systems, pp. 2203–2208 (2004)
9. Hirth, J., Schmitz, N., Berns, K.: Emotional architecture for the humanoid robot head ROMAN. In: IEEE International Conference on Robotics and Automation, pp. 2150–2155 (2007)
10. Watanabe, A., Ogino, M., Asada, M.: Mapping facial expression to internal states based on intuitive parenting. J Robot. Mechatron. **19**(3), 315–323 (2007)
11. Hashimoto, M., Yamano, M., Usui, T.: Effects of emotional synchronization in human-robot KANSEI communications. In: IEEE International Workshop on Robot and Human Interactive Communication, pp. 52–57 (2009)
12. Lim, A., Okuno, H.G.: Developing robot emotions through interaction with caregivers. In: Vallverdú, J. (ed.) Handbook of Research on Synthesizing Human Emotion in Intelligent Systems and Robotics, pp. 316–337 (2014). doi:10.4018/978-1-4666-7278-9.ch015
13. Asada, M., MacDorman, K.F., Ishiguro, H., Kuniyoshi, Y.: Cognitive developmental robotics as a new paradigm for the design of humanoid robots. Robot. Auton. Syst. **37**, 185–193 (2001)
14. Ekman, P.: Methods for measuring facial action. In: Scherer, K.R., Ekman, P. (eds.) Handbook of Methods in Nonverbal Behavior Research. Cambridge University Press, New York (1982)

15. Mehrabian, A., Russell, J.A.: An Approach to Environmental Psychology. MIT, Cambridge (1974)
16. Scherer, K., Schorr, A., Johnstone, T.: Appraisal Processes in Emotion: Theory, Methods, Research. Oxford University Press, Oxford (2001)
17. Smith, C.A., Lazarus, R.S.: Emotion and adaptation. In: John, O., Robins, R., Pervin, L. (eds.) Handbook of Personality: Theory and Research, pp. 609–637. Oxford University Press, Oxford (1991)
18. Plutchik, R.: The Emotions. University Press of America, Lanham (1991)
19. Russell, J.A., Mehrabian, A.: Evidence for a three-factor theory of emotions. J. Res. Pers. **11** (3), 273–294 (1977). doi:10.1016/0092-6566(77)90037-X
20. Bosse, T., Pontier, M., Treur, J.: A dynamical system modelling approach to Gross' model of emotion regulation. In: ICCM 2007: International Conference on Cognitive Modeling. Taylor & Francis/Psychology Press, Oxford (2007)
21. Gross, J.: Emotion regulation: affective, cognitive, and social consequences. Psychophysiology **39**, 281–291 (2002). doi:10.1017/S0048577201393198
22. Russel, J.A.: Reading emotions from and into faces: resurrecting a dimensional-contextual perspective. In: Russel, J.A., Fernandez-Dols, J.M. (eds.) The psychology of facial expression, pp. 295–320. Cambride University Press, New York (1997)
23. Panayiotou, G.: Emotional dimensions reflected in ratings of affective scripts. Pers. Individ. Differ. **44**(8), 1795–1806 (2008)
24. Bakker, I., Van der Voordt, T.J.M., Vink, P., de Boon, J.: Pleasure, arousal, dominance: Mehrabian and Russell revisited. Curr. Psychol. **33**(3), 405–421 (2014). doi:10.1007/s12144-014-9219-4
25. Plutchik, R.: Emotion: A Psychoevolutionary Synthesis. Harper & Row, New York (1980)
26. Lucey, P., Cohn, J.F., Kanade, T., Saragih, J., Ambadar, Z., Matthews, I.: The extended Cohn-Kanade dataset (CK+): a complete dataset for action unit and emotion specified expression. In: IEEE Workshop on CVPR for Human Communicative Behaviour Analysis (2010)

A Combined Feature Extraction Method for Automated Face Recognition in Classroom Environment

Md. Shafiqul Islam, Asif Mahmud, Azmina Akter Papeya,
Irin Sultana Onny, and Jia Uddin(✉)

BRAC University, 66 Bir Uttam AK Khandakar Road Mohakhali,
Dhaka 1212, Bangladesh
shafiqulislam561@gmail.com, typeamahmud@gmail.com,
azmina721@gmail.com, irinsultanaonny@gmail.com,
engrjiauddin@gmail.com

Abstract. Face recognition is a pattern recognition technique and one of the most important biometrics; it is used in a broad spectrum of applications. Classroom attendance management system is one of the applications. This paper proposes an optimized method of face detection using viola jones and face recognition using SURF and HOG feature extraction methods. The proposed model takes a video frame from an input device, then it detects faces in that frame using proposed optimized face detection method. Lastly, the detected faces are matched with pre-loaded customized database using proposed face recognition method. In addition we have tested our model with other existing model using two different customized datasets. Without human intervention this proposed model almost accurately completes the attendance of students in a class.

Keywords: Face detection · Face recognition · Classroom · Attendance · Kinect · SURF · HOG · Viola-Jones

1 Introduction

Biometrics authentication is used in computer science for various security purposes or to identify human [1]. There is different type of methods of biometric identification: face recognition, fingerprint identification, hand geometry biometrics, retina scan, iris scan, signature, voice analysis [2]. Among those methods face recognition has a distinct advantage due to it's non-contact process [3]. Conveying people's identity human face plays an important role [4]. Several applications can be built using face recognition for different purpose [5]. Classroom attendance management problem is one of them. As traditional attendance system is a manual process. Roll calling, card punching or paper based attendance is time consuming. Also manually recorded attendance can be easily manipulated [6]. To solve those entire problems this paper suggests an effective solution.

This paper introduces a system to solve traditional hectic classroom attendance management system. This system will mark attendance of those students who will be present in the classroom and registered for the class. We have experimented with

S.M. Thampi et al. (eds.), *Advances in Signal Processing and Intelligent Recognition Systems*, Advances in Intelligent Systems and Computing 678,
DOI 10.1007/978-3-319-67934-1_38

different feature extraction method systems to find an efficient feature extraction system. This system ensures accurate attendance, saves time, and convenience.

The rest of the paper is organized as follows: Sect. 2 includes literature review, Sect. 3 contains the proposed model, Sect. 4 consists of experimental setup and results analysis and lastly Sect. 5 contains the conclusion.

2 Literature Review

A number of works have done in this field. Visar Shehu et al. have proposed an automated attendance management system using computer vision algorithm [7]. For face recognition they have implemented Eigen face methodology. Eigen face has some problem such as the occurrence of class overlapping increases when more face classes are represented by the same face space, thus lowering the recognition rate [8]. In our work we deal with this problem and it gives proper result now. A problem faced during this process was the large number of false-positives which are the objects mistakenly detected as faces. While capturing the images students have to pay attention on the camera, which may interrupt the class regular environment. Samuel Lukas, Aditya Rama Mitra has proposed a student attendance system in classroom using face recognition technique to capture the students who are present in the class [9]. In order to recognize the face they have combined two methods which are Discrete Wavelet Transforms (DWT) and Discrete Cosine Transform (DCT). Priyanka Wagh, Jagruti Chaudhari has proposed an attendance System based on face recognition using eigen face and PCA Algorithms [10]. Jonathan Chin performed experiments named automated attendance capture and tracking system by placing webcam on the laptop to continuously capture the video of the students. Viola-Jones algorithm is used for face detection due to high efficiency. Eigen face methodology is used for face recognition. However, the students are required to remain alert as the Eigen face methodology is not capable of recognizing the titled faces captured in the frames. Also, a small classroom has been used due to the limited field of view of the webcam used on the laptop. Muhammad Fuzail, Hafiz Muhammad, and Fahad Nouman, presented a survey paper named Face Detection System for Attendance of Class Students uses HAAR classifier for face detection and Eigen Face methodology for face recognition [11]. This system still lacks the ability to identify each student present on class. Those are some related paper of our topic from where we took knowledge and idea to develop new version.

3 Proposed Model

Figure 1 shows the block diagram of our proposed model. Our model can be divided into two distinctive parts, face detection section and face recognition section. Firstly, the system takes input video frame from Kinect camera and human frontal face will be detected using viola jones algorithm. After detection of a face or more than one faces, the system first crops those faces from the video frame and coverts those into Gray scale images. Secondly non facial images and overlapping are removed, many viola jones algorithm sometimes detects non facial image as face. In this step we remove those non facial images from the detected facial image list. This step concludes the facial detection part. Next face recognition part begins. Hog and SURF features are detected and extracted from detected faces. On the other hand we have a database containing images of test persons. Currently we are using sixty images of each person. HoG and SURF features from every dataset images are detected, extracted, combined and fed into an SVM classifier. Lastly features of detected faces and test faces are compared in the classifier and the result is shown.

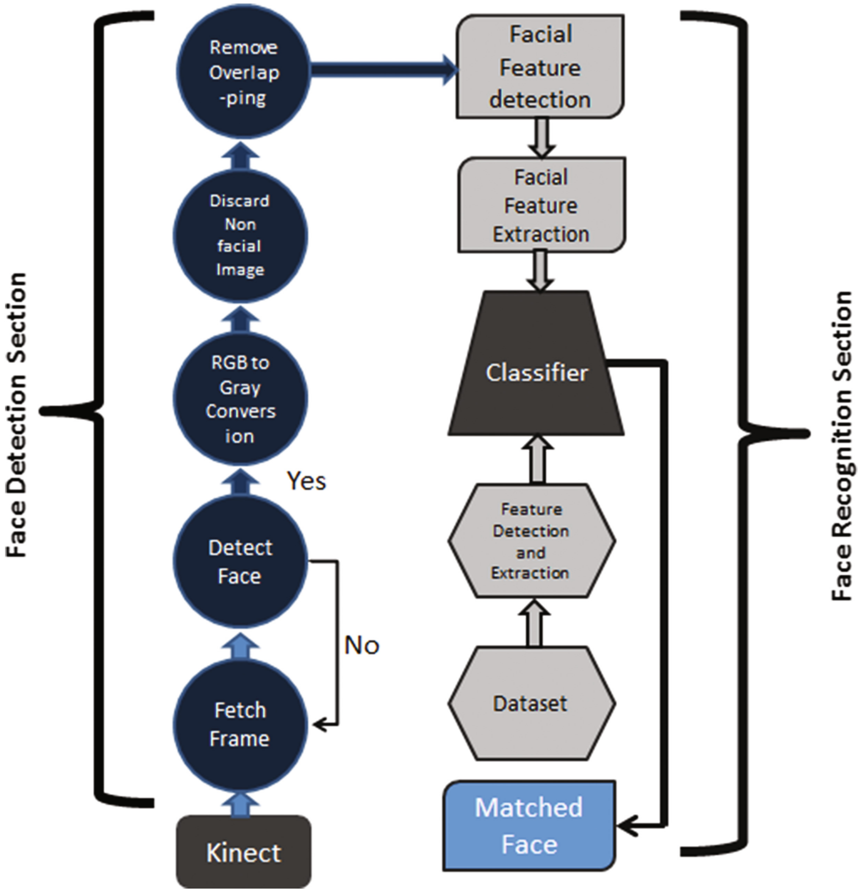


Fig. 1. Block diagram of proposed model

Moreover we have implemented a database to store the result of matched faces.

3.1 Face Detection

We have used viola jones algorithm for face detection as it is an effective way to detect face. However viola jones is not hundred percent accurate as in some cases many non-facial and overlapping elements are detected as face in this algorithm. We have tried to reduce the error percentage by implementing our developed method to reduce error percentage.

Viola Jones Algorithm Overview. Viola jones face detection algorithm is a widely used method for real-time object detection [12]. Viola jones method for detecting human face contains three techniques Integral Image, adaboost algorithm, cascade classifier.

Features and Integral Image. Viola Jones algorithm uses Haar like features. The integral image at location x,y contains the sum of the pixels above and to the left of x,y inclusive [13]. Where I(x,y) is the integral image and i(x,y) is the original image (in Fig. 2). Using the following pair of recurrences:

$$s(x,y) = s(x,y-1) + i(x,y) \tag{1}$$

$$I(x,y) = i(x-1,y) + s(x,y)$$

Where s(x, y) is the cumulative row sum, s(x, −1) = 0, and I(−1, y) = 0.

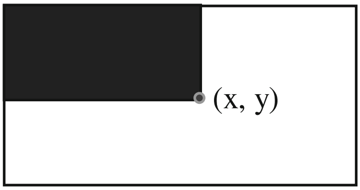


Fig. 2. Integral Image Point (x, y)

Adaboost algorithm. From the rectangle features available, an algorithm need to choose the features that give the best results. Viola Jones algorithm chose a variant of Adaboost to select features and to train a classifier [14]. Adaboost is a machine learning algorithm. It trains a set of weak classifiers to develop a strong linear classifier.

Originally:

$$hj(x) \in \{+1,-1\} \tag{2}$$

Form of Linear Combination:

$$C\theta(x) = \left(\sum\nolimits_t h_t(x) + b\right) \tag{3}$$

Cascade Classifier. Cascade of classifiers increased detection performance and reduce computation time [14]. Smaller, and therefore more efficient, classifiers can be constructed which reject many of the negative sub-windows while detecting almost all positive instances. The goal of each stage is to remove false faces.

RGB to GRAY scale conversion. Color increases the complexity of the model. Handling RGB color image is more complex than grayscale image. It is relatively easier to deal with a single color channel than multiple color channels [14]. Therefore gray scale images are used in our model.

Removal of Non-facial and Overlapping faces. As facial detection is a very costly process, resulting inaccuracy in this method can sometime lead the costs to a greater amount. Before removing non-facial and overlapping faces we have first down sampled the captured RGB frame and then has normalized it by using Gaussian filtering method in order to increase the efficiency of the facial detection. The default value of medianBoxsize is set to the size of the facial bounding box of the first frame and boxSizeThreshold is set to 100. The median takes the value of the biggest size of the box if there are multiple bounding boxes. As the method proceeds medianBoxsize is calculated by taking the current median value with the current 2D-array faceboundingBox for each captured frame. This fixes the size of the face bounding box. The median value is then checked whether it falls within a given range. This condition has been set in our method which ensures that final face bounding box will be always encircled with in the face region. Therfore, removing the overlapping faces. For to remove non-facial faces, frames that have only one bounding box applicant, the applicant becomes the final face bounding box. For image frames that have multiple facial bounding boxes, the candidate that has a box size within threshold value of the *medianBoxSize* is used as the final facial bounding box. Through our experimentation, using 100 pixels as the threshold gave the best result.

3.2 Face Recognition

Face recognition is an integral part of our proposed model. We have experimented with different feature extraction methods and combination of methods to find an efficient combination of features. We have experimented with HOG, SURF, LBP, Combination of HOG-SURF and Combination of HOG-SURF-LBP. Among all of these combinations of HOG-SURF returned the best result. More on this will be discussed in the experimental setup and result analysis section. In this part, HOG and SURF features from each dataset image is extracted, combined and stored in a 2D array. Then this array is fed into an SVM classifier. On the other side detected face's HOG and SURF features are also extracted. Lastly features of Dataset images and detected facial images are compared and best match is shown in the result.

Speeded Up Robust Features (SURF). SURF is a local feature detector and descriptor. SURF uses a hessian based blob detector to find interest points [15]. Given a point x = (x, y) in an image I, the Hessian matrix H(x,σ) in x at scale σ is defined as follows:

$$H(x,\sigma) = \begin{bmatrix} L_{xx}(x,\sigma) & L_{xy}(x,\sigma) \\ L_{xy}(x,\sigma) & L_{yy}(x,\sigma) \end{bmatrix} \tag{4}$$

Where $\mathrm{Lxx}(\mathrm{x},\sigma) = \mathrm{I}(\mathrm{x}) * \partial 2 \partial \mathrm{x} 2\mathrm{g}(\sigma)$, the convolution of the Gaussian second order derivation with the image I in point x and similarly for $\mathrm{Lxy}(\mathrm{x},\sigma)$ and $\mathrm{Lyy}(\mathrm{x},\sigma)$.

An Integral image I(x) is an image where each point x (x,y) stores the sum of all pixels in a rectangular area

$$I(x) = \sum_{i=0}^{i \leq x} \sum_{j=0}^{j \leq y} I(x, y) \tag{5}$$

Histogram of Oriented Gradients (HOG). HOG is a reliable feature extraction system mainly used in image processing for object detection. HOG works by dividing the image into very small connected regions which are called cells and for each cell, finding histogram of gradient direction inside the cell.

First step of calculation is the computation of the gradient values. The most common method is to apply the 1-D centered, point discrete derivative mask in one or both of the horizontal and vertical directions. Specifically, this method requires filtering the color or intensity data of the image with the following filter kernels:

$$[-1, 0, 1] \text{ and } [-1, 0, 1] \top. \tag{6}$$

4 Experimental Setup and Result Analysis

To evaluate the proposed model, we create our own dataset for the students. In fact we have created two different set of datasets. First one is a set of three persons consisting of sixty images per person which are used for the actual system. On the other hand the second dataset is a set of nine persons consisting of ten images per person. This second dataset is used to test different feature extraction methods and combinations of them to find the best output. Figure 3 shows the dataset of stored images.

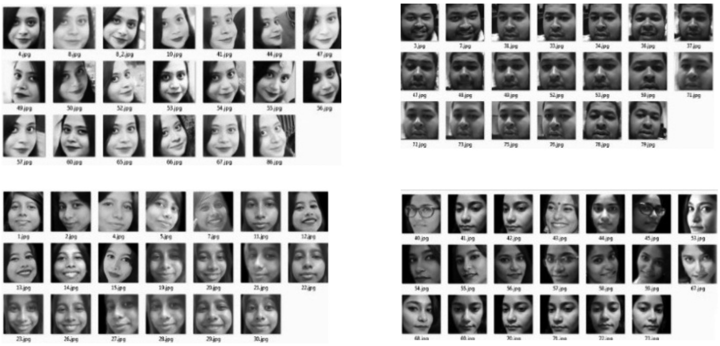


Fig. 3. Image of dataset

Dataset contains facial images of each person in different facing direction in order to get maximum variation. Dimension wise, each image is 200 × 200 pixels.

As a platform for our system we have used MATLAB R2016a. For to connect camera with our system we needed Image Acquisition Toolbox. Image Acquisition Toolbox provides functions and blocks that enable to connect cameras to MATLAB and Simulink.

We have used vision.cascadeObjectDetector library for using viola jones algorithm and fitecoc library for SVM classifier.

In Fig. 4, we can see that stock algorithm cannot detect all faces in the video frame. Also it is detecting a false face.

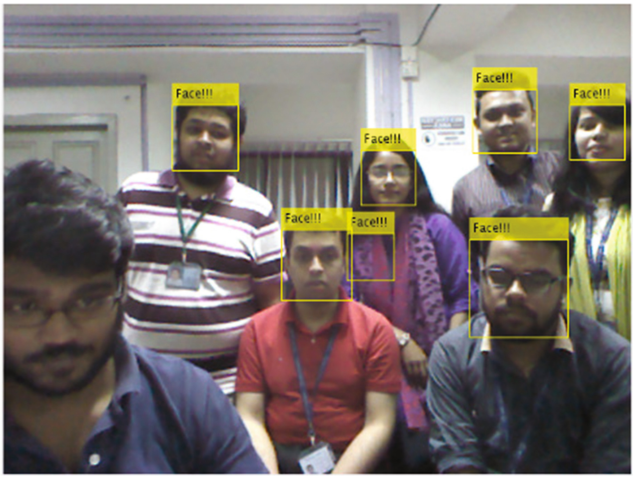


Fig. 4. Face detection with false face

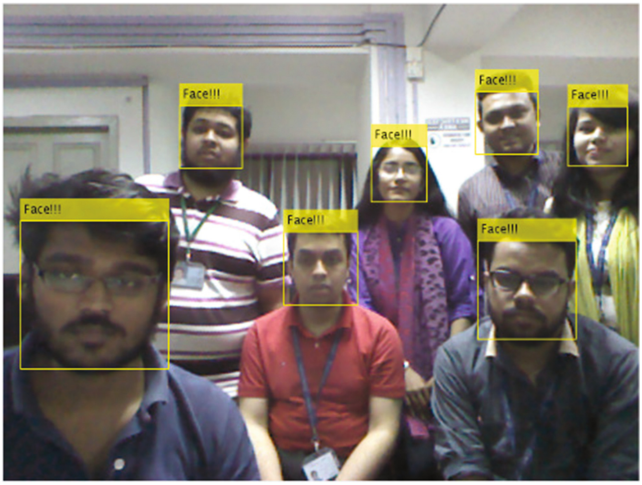


Fig. 5. Face detection with proper face

Figure 5 shows the difference between stock algorithm and our modified algorithm. We can see that no false face has been detected and it can detect all faces present in the frame.

The following Table 1 shows the representation of our work. The percentage of efficiency has been calculated by total number of faces detected/total number of persons present in the environment. The table also shows the result of removing non-facial and overlapping faces by our system.

Table 1. Result of detected face and comparison.

	# of person	# of detected face	# of false face	Efficiency	Overlapping
Built-in Function	7	6	1	85%	1
Used Code	7	7	0	100%	0

As stated earlier, we have experimented with different feature extraction methods and combination of more than one methods to find the optimum solution. We have used the second dataset as a testbed. Finally we have found that a combination of HOG and SURF feature extraction method can provide best solution. We have used the second dataset and test pictures for this comparison. We have extracted 20736 feature points for each pictures using HOG, 12800 feature points for each pictures using SURF feature extraction method. Combining HOG and SURF feature extraction methods, we got 33536 feature points. We have also tested LBP feature extraction method but this method is really not suitable for face recognition applications. So we've omitted LBP. Tabular representation of the result is shown in Table 2. 9 person's dataset image and test image has been used for producing this result.

Table 2. Comparison between HOG, SURF and LBP for matching.

Name of Algorithm	# of Persons	# of match faces
HOG	9	6
SURF	9	1
HOG + SURF	9	7

We have also tested the feature method combination's efficiency in video frame scenario as well. For this situation, we have used the first dataset. Same dataset, background and test person have been used for this experiment. The combination of HOG and SURF also proved efficient in this scenario as well. Tabular representation shown in Table 3. The results are represented in invert formation i.e. more positive value means better feature extraction method.

Table 3. Matching percentage of HOG, HOG + SURF, HOG + LBP.

Algorithm	Percentage
HOG	−0.2046
HOG + SURF	−0.1990
HOG + LBP	−0.2046

*Note: the results are shown in invert formation.

5 Conclusion

The proposed method is to result to an optimal face detector by minimizing the inaccuracy often caused by Viola Jones facial detection method, and a better facial recognizer using HOG, SURF features for SVM Classifier. This paper also shows the comparison between HOG, SURF, HOG + SURF for better feature extraction from video frame image and matching percentage between HOG, HOG + SURF, HOG + LBP. All the testes are done in a real time attendance system which has been made by us. Sometimes there is a problem with matching, in future we will work on the accuracy of the recognition part.

References

1. Hassan, M.: Biometric Industry Year-End Review 2016, M2SYS Blog On Biometric Technology, 02 January 2017. http://www.m2sys.com/blog/comments-on-recent-biometric-news-stories/2016-biometric-industry-year-end-review/. Accessed 22 Mar 2017
2. Biometric authentication: what method works best?, Biometric authentication: what method works best? http://www.technovelgy.com/ct/Technology-Article.asp?ArtNum=16. Accessed 22 Mar 2017
3. NEC, Face Recognition, Face Recognition: Technologies: Biometrics: Solutions & Services | NEC. http://www.nec.com/en/global/solutions/biometrics/technologies/face_recognition.html. Accessed 22 Mar 2017
4. Bhadauria, A., Goel, M., Mehta, S.: Biometics: face recognition system & applications. Int. J. Sci. Res. Eng. Technol. **1**, 138–140 (2014)
5. Andrew W., Bolle, R.M.: Face recognition and its applications, IBM T.J. Watson Research Center. http://www.andrewsenior.com/papers/SeniorB02FaceChap.pdf
6. Trader, J.: The Value of biometrics for student attendance management systems, M2SYS Blog On Biometric Technology, 07 April 2016. http://www.m2sys.com/blog/education/the-value-of-biometrics-for-student-attendance-management-systems/. Accessed 22 Mar 2017
7. Shehu, V., Dika, A.: Using real time computer vision algorithms in automatic attendance management systems. In: 32nd International Conference on Information Technology Interfaces (ITI), pp. 397–402, June 2010
8. Deschamps, M.: Advantages and limitations (2014). http://mathdesc.fr/, http://mathdesc.fr/documents/facerecog/AdvantagesLimitations.htm. Accessed 6 Mar 2017
9. Lukas, S., Mitra, A.R., Desanti, R.I., Krisnadi, D.: Student attendance system in classroom using face recognition technique. In: International Conference on Information and Communication Technology Convergence (ICTC) (2016)

10. Wagh, P., Thakare, R., Chaudhari, J., Patil, S.: Attendance system based on face recognition using eigen face and PCA algorithms. In: International Conference on Green Computing and Internet of Things (ICGCIoT) (2015)
11. Fuzail, M., Muhammad, H., Nouman, F.: Face detection system for attendance of class students. Int. J. Multi. Sci. Eng. **5**(4), 6–10 (2014)
12. Young, M.: The Technical Writer's Handbook. University Science, Mill Valley, CA (1989). Viola–Jones object detection framework, Wikipedia. 01-Mar-2017
13. Viola, P., Jones, M.J.: Robust real-time face detection. Int. J. Comput. Vision **57**, 137–154 (2004)
14. Parekh, N.: In image processing applications, why do we convert from RGB to Grayscale? Quora, 28 July 2016. https://www.quora.com/In-image-processing-applications-why-do-we-convert-from-RGB-to-Grayscale. Accessed 22 Mar 2017
15. Bay, H., Ess, A., Tuytelaars, T., Gool, L.V.: Speeded-Up Robust Features (SURF). Comput. Vis. Image Underst. **110**(3), 346–359 (2008)

A Low Cost Intelligent Smart System for Real Time Child Monitoring in the School

Pruthvi Raj Myakala[1(✉)], Rajasree Nalumachu[1], and V.K. Mittal[2]

[1] Indian Institute of Information Technology Chittoor, Sri City, India
{pruthvi.m14,rajasree.n14}@iiits.in
[2] Ritwik Software Technologies Pvt. Ltd., Hyderabad, India
DrVinayKrMittal@gmail.com

Abstract. Speech processing techniques help improving real-time communication between human to human, human and machine, machine to machine. If these techniques are used integrated with robots, then their physical flexibility and wider reach can enable a wide range of real-time applications. In this paper, we propose an *'Intelligent Cry Detection Robotic System' (ICDRS)* for real-monitoring of child-beating in classrooms, in order to facilitate prevention of child-abuse prevalent in this form. The proposed system has two major modules: the 'Cry-Detection System' (CDS) and a 'Smart Robotic System' (SRS) equipped with audio-visual sensors. The CDS unit present in the classroom, consist of three parts. First, the cry-recording unit (CRU) records the audio signals and sends to 'Signal Processing Unit' (SPU). Then SPU applies signal processing techniques, and intelligently detects the cry events using the features extracted from the acoustic signal. If the system detects a cry, then it further sends the control commands to the 'Signal Transmission Unit' (STU), which sends an automatic SMS to the Vice-Principal or Supervisor-Teacher, i.e., the person-in-charge, thereby alerting him/her about the child cry in a particular classroom. The controlling person can give control-commands to the SRS from a web-application and can get the live-stream of the video from the classroom. A Wi-Fi module acts to facilitate the communication between this controller and the Robot (SRS). The initial performance evaluation results are very much encouraging. The proposed system can have potential applications in the schools, hospitals and child care centers etc. Hopefully, this prototype can be a useful step towards preventing child-abuse, prevalent in different forms in our society.

Keywords: Cry detection robot · Energy · Cry signal · Cry analysis · WiFi controlled · Remote control operation

1 Introduction

Generally robots are highly sophisticated and powerful machines which are designed to perform tedious work In addition, they assist the human beings

S.M. Thampi et al. (eds.), *Advances in Signal Processing and Intelligent Recognition Systems*, Advances in Intelligent Systems and Computing 678,
DOI 10.1007/978-3-319-67934-1_39

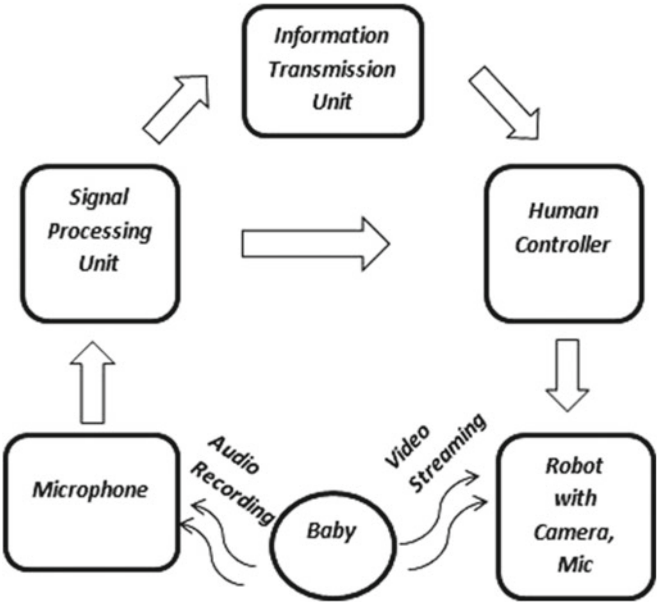


Fig. 1. Architecture of the *ICDRS* which consists of *Cry Detection System* and *Smart Robotic System*

in many aspects of the life. Initial robots were developed in 1940s. They were mainly used for handling the radioactive materials and then in late 1960s these were also used for picking and placing of the objects in different places. Connectivity between the controller and the robot can be through some wired connection or through wireless communication [1]. With advancement of technology connectivity has shifted to wireless communication because of very less range of control operation of wired networks. Through wireless Communication range of control can be increased. Nevertheless, it is also limited. Wireless Communication can take place using different technologies like Bluetooth, WiFi, RF, IR etc.

On the other hand, speech is one of the easiest ways of communication between people for expressing views, ideas and thoughts. Speech Signals can be classified into voiced [2] and unvoiced sounds. Voiced speech signals are because of the vibration of the vocal folds, for example these would be vowels. Unvoiced speech signals don't involve the vocal folds vibrations, for example these would be fricative sounds [3]. In the same way Cry, Shout, Laughter also falls under the category of paralinguistic speech signals, where these kind of signals consists of some voicing activity in them.

Now a days child abuse and beating of children is getting more predominant in every part of the world especially in schools. Sometimes, higher authorities may not know about the issue and may not be able control situation, which can lead to any dangerous situation. In order to handle such issues, most of the high end schools have installed surveillant cameras in the classrooms for monitoring the students, However deploying these cameras in the every classroom is costly and it also requires an human to watch the surveillance footage every moment in order to monitor the students in the classrooms which is difficult task. This paper proposes an *Intelligent Cry Detection Robotic System (ICDRS)* in order to address the above mentioned problem, This automated system will monitor the classroom, However the *Cry Detection System* deployed in the classroom will detect the cry and notify the controller with an alert message.

In this paper, we developed an *ICDRS* for classroom monitoring. This integrates speech signal processing techniques and wifi controlled robot for this application. It processes cry signal (paralinguistic speech signal) [4] and extracts the energy feature in the *CDS*. Signal energy is the combined feature of source and system [5]. The developed prototype validates the cry using this combined feature i.e., Energy of the signal for detecting and analyzing the cry.

ICDRS is divided into two parts, one is the *Robotic system* and another is the *Cry detecting system*. The *Robotic system* can be remotely operated by the controller using WiFi communication medium. *Robotic system* is controlled by the controller and it travels to the classrooms according to the control commands it receives. This will live stream the video of the classroom with the camera mounted on it. Before the *robotic system* starts its operation, the *Cry Detection system* will be recording the sound signals and analyzing them on the microcontroller, if the *cry detection system* finds any cry in the classroom it will notify the controller with an SMS saying that child is crying in the classroom with the help of the GSM module integrated in the system.

This paper is organized as follows. In Sect. 2, it gives an overview of the design aspects of the *Intelligent Cry Detection Robotic System (ICDRS)* in both hardware and the software aspects for developing the *Cry detection System* and the *Robotic platform*. In Sect. 3, Intelligent Detection and Control operation, Cry detection, web based control and also the video/image streaming are discussed. Section 4, gives the performance evaluation which includes the various results and experiments carried out in the *ICDRS*. Explanation about the advantages of the prototype are given in the Sect. 5. Possible applications are discussed in detail in the Sect. 6 and followed by summary and future scope in the Sect. 7.

2 Design Details of *Intelligent Cry Detection Robotic System (ICDRS)*

This section explains about the hardware and software architectures followed in building the *ICDRS*. Figure 1, explains about the functional block diagram of the *ICDRS*. Figure 2, gives the overall working architectural design of the *ICDRS* prototype which includes all the key components being used. The *robotic system* block explains how the robot is controlled using WiFi connecting medium and also the cry recording, processing [6] and generation of SMS is explained in the *cry detection system* block.

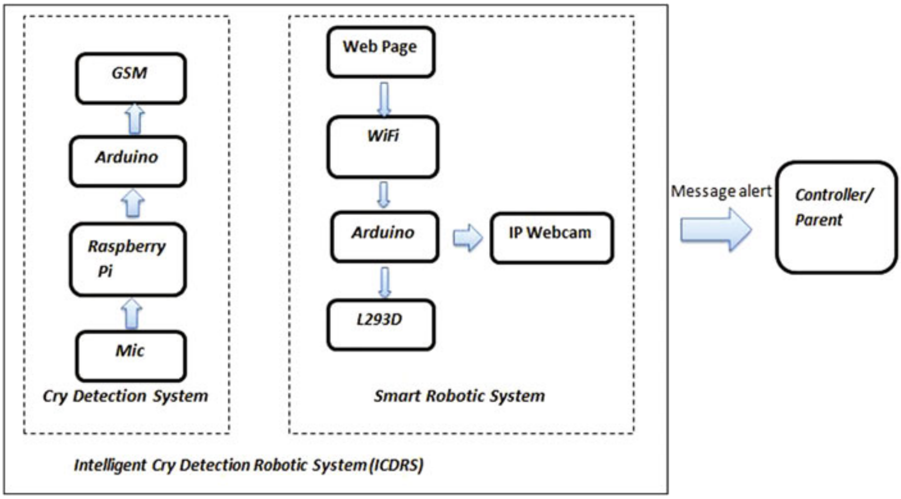


Fig. 2. Architecture of the *ICDRS* which consists of *Cry Detection System* and *Smart Robotic System*

2.1 Design Details of the Smartbot Prototype

The key components used in building this part are Arduino an 8-bit microcontroller with a clock speed of 16 MHz, ESP8266 is an low cost WiFi module for the connecting medium, which as a specialty of acting as both client and server. L293D, A dual H- bridge driver integrated circuit is used to amplify the low current control signal from the microcontroller and produce a high current control signal for the DC motors movement. DC motors are fixed to the robotic chases [7] and pins of the motors are connected to the motor driver integrated chip for receiving the control commands from the microcontroller [8]. Table 2, gives all the necessary AT commands for configuration of the WiFi module to network [9].

Table 1. AT Commands of GPRS/GSM for sending SMS from *Cry Detection System* to the Person in-charge

(a) AT command	(b) Meaning
ATD	Dial
AT+CMGS	Send message
AT+CMMS	More messages to send
AT+CMGR	Read SMS message
AT+CMGC	Send command
AT+IPR=0	To choose auto baud rate

Table 2. AT Commands to configure ESP8266 for controlling the Robotic System by the Person in-charge

(a) AT command	(b) Meaning
AT+RST	Reset command
AT+GMR	To check Firmware version
AT+CWMODE?	To check mode of operation
AT+CWLAP	To check available Wi-Fi networks
AT+CIFSR	To check IP address
AT+CWJAP	To connect to suitable Wi-Fi

2.2 Design Details of the *Cry Detection System (CDS)*

The key components used in building the *CDS* are Raspberry pi 3, Model B, which is the heart of the entire system. Matlab Simulink models are deployed into the microcontroller and processed in it [10]. Zebronica microphone is used to record the sound. Another end of the microphone is connected to the 7 channel USB soundcard and further connected to the microcontroller. After processing an control signal is sent to the arduino which is connected to the raspberry pi for the further process. Arduino will communicate with the GSM module and send an SMS to the controller. GSM module used in the system is sim900a, its on the quad band technology which includes bands of 850/900/1800/1900 MHz. AT commands are used for the interaction of the GSM with arduino. Sending an SMS to the controller based on the child cry to the pre-configured number is done using the AT commands. Table 1, consists of all required AT commands for configuration and sending SMS for the GSM module [11].

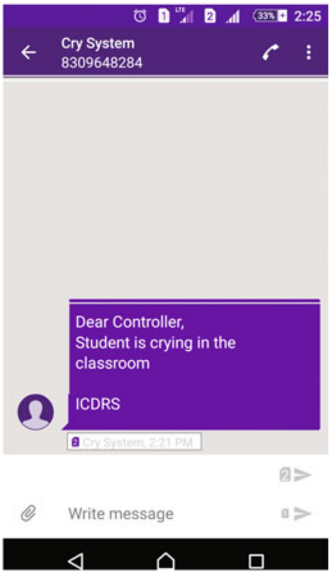
Initially it record the signal and processed in the matlab Simulink model. MATLAB Simulink [12] is a simulation tool, it is a design environment for embedded systems which are integrated with the MATLAB. This enables to compute MATLAB algorithms. Further this also encourages to support system level design, simulation, code generation and testing of the embedded systems.

For detecting the Cry signal of the child, Combined feature of source and system i.e., Energy of the signal is used in the *ICDRS*. Energy of the signal is the summation of energy across all the frequency components of signal spectral energy density. For *ICDRS*, energy is calculated using the auto correlation [13]. Highest peak after applying the auto correlation is the energy of the signal, when the time-lag(m) is zero.

3 Intelligent Detection and Control Operation

3.1 Cry Detection

Cry detection system makes *ICDRS* as a real time system by processing real time information. Mic present in the system continuously records the audio signals and



Dear Controller,
Student is crying in the
classroom

ICDRS

Fig. 3. Screen shot of the SMS received by the Vice-Principal/Supervisor when his/her ward is crying

sends them to the processing unit in buffers. Processing unit analyses this data by extracting features from the data. These features are extracted by applying signal processing techniques on them. This work is done by the SIMULINK model dumped in the processing unit. Later, the values obtained are validated against the data present in the unit, which helps the system to intelligently identify the cry signal. As soon as the unit identifies the cry signal it sends control command to the microcontroller intimating it about the baby cry. This, in turn sends message to the controller, alerting him/her about the child cry in the classroom (Fig. 3).

3.2 Web Based Control

WiFi module i.e., ESP8266 doesnt require any serial communication using wires, it gives a flexibility of wireless [14]. The control commands are sent from the web page with required instructions, those instructions are generated in the form of a HTTP request and it is received by the WiFi module which is mounted on the Smartbot System. Those signals are decoded and processed for the smartbot movement. For movements, forward, backward, left, right directions are written on the web page. Depending on the button operated by the controller on the web page the robots moves in that direction [15]. In the same way the video streaming [16] is also done by clicking of the button the page [17].

3.3 Video Streaming/Images of the Classroom

In order to conform that child is crying in the classroom, the smartbot is sent to the classroom to capture the image or video. This also confirms that if there is any conflict going on in the class for the child cry. In this paper video streaming [18] done using the Internet Protocol(IP) Camera which is compatible with the WiFi [19]. This camera is mounted on the smartbot based on the control command [20] the camera module turns on capture the real time information of the classroom [21]. The captured images or videos can be seen on the smart phone or in the web browser [22]. Figure 2, shows the screen shot of the SMS received by the controller from the *ICDRS*.

4 Performance Evaluation

Various delays encountered during the operation of the system are tabulated in Table 3. This table consists of an average values of delay time, when the experimentation is carried out for 50 times. It is observed that if network is not good near the ICDRs, then the SMS will not be sent by the system to the controller/parent, because GPRS used in the system requires good network connectivity for the SIM present in the module to communicate to the GPRS module. Time taken for the SMS to transmit to the controller depends on the connectivity of the network.

Table 3. *Intelligent Cry Detection Robotic System (ICDRS)* response time for various control commands

(a) Control commands propagation	(b) Response
Sending SMS	1.2 s
Video Streaming	2 s
Control commands to ESP8266	1.01 s
Signal Processing in Raspberry Pi	0.2 s

In addition time taken for the control commands to execute after selecting a button on the web page is about 1.01 s. This is the time taken by WiFi module to send the information to the micro-controller present on the robotic platform [23]. Time taken for the processing of signals is about 0.2 s, which is less compared to all other operations. This may be differed when, size of the data analyzed changes, probably it increases for large data and decreases for small data size. Delay time encountered for live video streaming is on average 2 s, which is more compared to all other operations.

5 Advantages of *ICDRS*

5.1 Intelligent Smart System

Cry Detection system present in the *ICDRS* analyzes audio signal, extracts features from it and validates it with the features of the cry signal and decides whether the signal is a cry signal or not.

5.2 Real Time Monitoring

Mic present in the *cry Detection system* records the audio signals and processes them in real time. In addition, it sends the message to the controller regarding the child cry and alerts him about the situation. Controller gives the control commands to the robot through Wi-Fi [24] and can see the live stream of the video of the classroom [25]. All these operations take place in real time, so the proposed system is a real time monitoring system.

5.3 Future Extension Work

This project is further extended to analyze the infant babies for real time monitoring. This would be extended to monitor the environment and climatic changes around the baby cradle. As a future progress reason for the cry would be analyzed. As previous work [11] helped to classify the baby cry into various categories like pain, discomfort, hungry etc.

6 Possible Applications

The developed *Intelligent Cry Detection Robotic System* has wide area of applications in day to day life, which is an real-time application of the speech signal processing in the world where this could be built with an low budget. Few important key applications are discussed in this section.

6.1 Prevention of Child Abuse

Physical maltreatment of children is present in all corners of the world, it disturbs the children mentally and physically. In some nations it is even considered as a crime. In most of the scenarios children may not be able to convey their pain to their parents, may be because of fear or any other reason. This case is more in children below age of 5. The proposed *ICDRS* can be modified to solve these kind of issues, where the system detects the child cry and tries to find the reason behind the cry, identify the maltreatment like beating scenario and notify the parents or to the respective people to aware them about the situation for further reaction.

6.2 Baby Care Centers

Most of the parents are forced to leave their children in baby care centers or with a nanny or with a relative in their absence, because of their tight scheduled work. But, they are panic and are curious about their child, they don't have any idea whether the person with whom they left their baby are able to take care of their child. In addition they always worry whether their child is crying or is he/she fine. The proposed system can to help parents to get free from their worries. The system can be modeled to monitor babies and detect the child cry and alert the parent [26]. It can even analyze the cry of the baby and find the reason behind their cry, which helps the parents and the care takers to take care of the baby accordingly. System can detect the environmental changes around the baby and notify the parent about the situation, it can also try to calm down the baby by playing light music or blinking different lights [27].

6.3 Applications to Study of Other Paralinguistic Sounds

Cry signal is one of the paralinguistic sound, laughter, cough and shout also falls under this category. The proposed system can be modified and used for applications of other paralinguistic sounds. The work on analysis of cough [4], laughter [28], shout [29] has been done earlier where various features like Energy and Formant frequencies are studied, the analysis on these paralinguistic features can be embedded with the hardware and systems for various applications could be developed.

6.4 Infant Cry Analysis

Infant cry as it is also an audio signal carries some meaningful information, which can be analyzed by using signal processing techniques [30]. The information obtained from the features extracted can be used to identify the reason for baby cry [31], that is analyzing the cause of cry like pain [32], discomfort, anxiety, hunger etc., [33] Analysis can be helpful to identify the baby who is crying from among many babies in a child care center or a play school [34]. In addition we can identify the severity of baby cry and caution the parent or the respective person [35].

7 Summary and Conclusion

An *Intelligent Cry Detection Robotic System (ICDRS)* for classroom monitoring and real time information is developed in this paper. This system is classified into two parts smart robotic monitoring and *Cry detection system*. This *robotic system* can be controlled remotely using the WiFi connectivity using the ESP8266 (WiFi module) mounted on the robot which gives a flexibility of wireless. This paper also explains how the signal processing is embedded into the microcontroller for the cry detection. The developed system will record the signals and process them in the microcontroller. If the system encounters the cry then it would trigger the arduino and generate an SMS to pre-defined number using the GSM module to the concerned people.

The biggest drawback faced is to configuring the WiFi module before processing. For configuring it requires another router. Another limitation in the system is GSM module can send the SMS only when proper network is present near the system. The paper also provide comprehensive view of the embedding the MATLAB Simulink into the microcontroller and the WiFi technologies for the connectivity between the robot and the controller. This research could help the upcoming researchers for conducting tests and experiments.

References

1. Asthana, S., Varma, N., Mittal, V.K.: An investigation into classification of infant cries using modified signal processing methods. In: 2nd International Conference on Signal Processing and Integrated Networks (SPIN), February 2015, pp. 679-684 (2015)
2. Rabiner, L.: On the use of auto-correlation analysis for pitch detection. IEEE Trans. Acoust. Speech Sig. Process. **25**, 24–33 (1977)
3. Oppenheim, A.V., Schafer, R.W., Buck, J.R.: Discrete Time Signal Processing. Prentice-Hall, Englewood Cliffs (1989)
4. Mittal, V.K., Yegnanarayana, B.: Study of characteristics of aperiodicity in Noh voices. J. Acoust. Soc. Am. (JASA) **137**(6), 3411–3421 (2012)
5. Titze, I.R., Story, B.H.: Acoustic interactions of the voice source with the lower vocal tract. J. Acoust. Soc. Am. (JASA) **101**, 2234–2243 (1997)
6. Asthana, S., Varma, N., Mittal, V.K.: Preliminary analysis of causes of infant cry. In: IEEE International Symposium on Signal Processing and Information Technology, ISSPIT, 15–17 December 2014, pp. 468–473 (2014)
7. Kumar, K., Nandan, S., Mishra, A., Kumar, K., Mittal, V.K.: Voice-controlled object tracking smart robot. In: Proceedings of IEEE 3rd International Conference on Signal Processing, Computing and Control 2015 (ISPCC 2015), JUIT, Waknaghat, India (2015)
8. Sharma, S., Asthana, S., Mittal, V.K.: A database of infant cry sounds to study the likely cause of cry. In: 12th International Conference on Natural Language Processing (ICON 2015), IIITM-K, Trivandrum, India (2015)
9. Ryu, D., Kang, S., Kim, M., Song, J.-B.: Multi-modal user interface for teleoperation of ROBHAZ-DT2 field robot system. In: 2004 Proceedings of IEEEIRSJ International Conference on Intelligent Robots and Systems (IROS 2004), vol. 1, pp. 168–173. IEEE (2004)
10. Raj, P., Rajasree, N., Jayasri, T., Mittal, Y., Mittal, V.K.: A Web Based Intelligent Spybot. In: Proceedings of IEEE 3rd International Mining Intelligence and Knowledge Exploration (MIKE 2015), vol. 9648, pp. 472–481. IIIT Hyderabad, December 2015
11. Mittal, V.K.: Discriminating the infant cry sounds due to pain vs. discomfort towards assisted clinical diagnosis. In: 7th Workshop on Speech and Language Processing for Assistive Technologies, SLPAT 2016, San Francisco, USA, 13 September 2016, pp. 37–42 (2016)
12. MATLAB Simulink Projects. https://in.mathworks.com/discovery/simulink-projects.html. Last Viewed 1 June 2017

13. Mittal, V.K.: Discriminating features of infant cry acoustic signal towards automated diagnosis of cause of crying. In: 10th International Symposium on Chinese Spoken Language Processing (ISCSLP 2016), Tianjin, 17–20 October 2016
14. Morshed, N.M., Muid-Ur-Rahman, G.M., Karim, M.R., Zaman, H.U.: Microcontroller based home automation system using Bluetooth, GSM, Wi-Fi and DTMF. In: International Conference on Advances in Electrical Engineering (ICAEE) (2009)
15. Narvaez, L., Llanes, E., Hernandez, C., Poot, R., Chi, V.: Design and implementation of a system for wireless control of a robot. Int. J. Comput. Sci. Issues **7**, 191–197 (2010)
16. Dayma, D., Chavan, B., Kale, S., Tarle, B.S.: Smart Spy Robot. Int. J. Sci. Technol. Manage. **04**(02), 93–95 (2015)
17. Narvaez, L., Lanes, E., Hermandez, C., Poot, R., Chi, V.: Design and implementation of a system for wireless control of a robot. Int. J. Comput. Sci. Issues **7**, 191–197 (2010)
18. Makula, P., Mishra, A., Kumar, A., Karan, K., Mittal, V.K.: Voice-controlled object tracking smart robot. In: Proceedings of IEEE 3rd International Conference on Electronics, Computing and Communication Technologies (CONNECT 2015). IIIT Bangalore, India, 10–11 July 2015
19. Patoliya, J., Mehta, H., Patel, H.: Arduino controlled war feild spy robot using night vision wireless camera and android application. In: 5th Nirma University International Conference on Engineering (NUiCONE), pp. 1–5
20. Vaishnav, P., Tiwari, S.: Accelerometer based hand gesture controlled robot. Int. J. Sci. Res. (IJSR) **4**(3), 223–226 (2015)
21. Balakrishnan, M.: A smart Spy robot charged and controlled by wireless systems. In: IEEE Sponsored 2nd International Conference on Innovations in Information Embedded and Communication Systems, ICIIECS 2015, Coimbatore, 19 March 2015
22. Maheshwari, T., Kumar, U., Nagpal, C., Ojha, C., Mittal, V.K.: Capturing the spied image-video data using a FlexiControlled Spy-robot. In: 2015 Third International Conference on Image Information Processing, pp. 330–335, December 2015
23. Benefits of Robots - RoboWorx .https://www.robots.com/articles/viewing/benefits-of-robots. Last Viewed 15 May 2017
24. Mittal, Y., Toshniwal, P., Sharma, S., Singhal, D., Gupta, R., Mittal, V.K.: A voice-controlled multi-functional smart home automation system. In: Proceedings of IEEE 12th International INDICON Conference, Jamia Millia Islamia, New Delhi, India (2015)
25. Jain, P., Firke, P.N., Patil, T.S., Rode, S.S., Kapadnis, K.N., Kapadnis, K.N., et al.: RF based Spy robot. Int. J. Eng. Res. Appl. **4**(4), 06–09 (2014). (Version 2)
26. Lavner, Y., Cohen, R., Ruinskiy, D., Ijzerman, H.: Baby cry detection in domestic environment using deep learning. In: 2016 IEEE International Conference on the Science of Electrical Engineering (ICSEE), pp. 1–5, November 2016
27. Skogsdal, Y., Eriksson, M., Schollin, J.: Analgesia in newborns given oral glucose. Acta Paediatrica **86**(2), 217–220 (1997). http://dx.doi.org/10.1111/j.1651-2227.1997.tb08872.x
28. Mittal, V.K., Yegnanarayana, B.: Analysis of production characteristics of laughter. Comput. Speech Lang. **30**(1), 99–115 (2015)
29. Mittal, V.K., Yegnanarayana, B.: Effect of glottal dynamics in the production of shouted speech. J. Acoust. Soc. Am. **133**(5), 3050–3061 (2013)

30. Petroni, M., Malowany, M., Johnston, C., Stevens, B.: A new, robust vocal fundamental frequency (F0) determination method for the analysis of infant cries. In: 1994 Proceedings of IEEE Seventh Symposium on Computer-Based Medical Systems, pp. 223–228 (1994)
31. Chandralingam, S., Anjaneyulu, T., Satyanarayana, K.: Estimation of fundamental and formant frequencies of infants cries; a study of infants with congenital heart disorder. Indian J. Comput. Sci. Eng. **3**(4), 574–582 (2012)
32. Daga, R.P., Panditrao, A.M.: Acoustical analysis of pain cries in neonates: fundamental frequency. IJCA Spec. Issue Electron., Inf. Commun. Eng. ICEICE **3**, 18–21 (2011)
33. Mima, Y., Arakawa, K.: Cause estimation of younger babies' cries from the frequency analyses of the voice - classification of hunger, sleepiness, and discomfort. In: 2006 International Symposium on Intelligent Signal Processing and Communications, ISPACS 2006, pp. 29–32, December 2006
34. Neustein, A. (ed.): Advances in Speech Recognition: Mobile Environments, Call Centers and Clinics. Springer, New York (2010)
35. Cohen, R., Lavner, Y.: Infant cry analysis and detection. In: IEEE 27th Convention of Electrical & Electronics Engineers in Israel (IEEEI 2012), pp. 1–5. IEEE (2012)

Optimized Cost-Based Biomedical Workflow Scheduling Algorithm in Cloud

N. Mohanapriya(✉) and G. Kousalya

Department of Computer Science and Engineering, Coimbatore Institute of Technology, Coimbatore, India
mohanapriyan08@gmail.com

Abstract. Owing to the data deluge of biomedical workflow applications, researchers consider cloud as a promising environment for deploying biomedical workflow applications. As the Workflow applications consist of precedence-constrained tasks, it requires high computing resources for its execution. Scheduling of pertinent resource to biomedical workflow applications is an appealing research area. The key concern is that the workflow applications should be scheduled with the appropriate resource such that the overall execution time and cost would be minimized and correspondingly resource utilization is maximized. The proposed Optimized Cost Scheduling Algorithm (OCSA) addresses this issue by scheduling the workflows to a resource in such a way that it efficiently reduces the time and cost. The proposed OCSA algorithm is simulated rigorously in WorkflowSim on real biomedical workflow application and the results are compared with the existing workflow scheduling approaches in terms of cost and time. The simulation result shows that the proposed scheduling algorithm appreciably reduces the execution time and cost than the existing scheduling algorithms.

Keywords: Biomedical workflow scheduling · Cloud scheduling · Cost based scheduling · Workflows in the cloud

1 Introduction

Biomedical workflow applications consist of numerous interdependent tasks, and there exist huge data transfer between the tasks. These applications are generally represented as Directed Acyclic Graphs (DAGs), which brings a problem of resource scheduling in distributed systems. Cloud is a distributive computing environment that dynamically delivers scalable, on demand services through virtualization of hardware and software over the internet. Cloud is based on a market-oriented paradigm, where the services consumed by the customers are charged on pay-as-you-go model [1, 2]. The prospect of running workflow applications through the cloud is made attractive by its benefits. The essential benefits includes,

- Virtualization - Cloud gives the illusion of unlimited resources and this allows the user to acquire sufficient resources at any time.
- Elasticity - Cloud providers offer scalable resources to its users so that the resources are gained and released as per the requirements.

S.M. Thampi et al. (eds.), *Advances in Signal Processing and Intelligent Recognition Systems*, Advances in Intelligent Systems and Computing 678,
DOI 10.1007/978-3-319-67934-1_40

A notion of discovering the suitable resource from the heterogeneous resource pool for the workflow task is referred as scheduling [3]. The mapping of tasks to the resources is an NP-Complete problem [4]. Scheduling of biomedical workflow applications involves substantial communicational and computational costs, which strongly emphasizes the usage of cloud computing for their execution. Cloud providers offer heterogeneous computing resources with different capabilities at various prices. Generally, high computing resources are expensive than the slower resource. Hence different scheduling is possible for the same workflow, which in turn impacts the scheduling time and cost. Therefore a special care for scheduling should be taken to avoid the unnecessary cost.

An optimized cost based workflow scheduling algorithm is proposed to schedule the biomedical workflow application in the cloud to minimize the overall execution time and cost for the execution of the workflow.

2 Related Works

Suraj pandey et al. [5] proposed the particle swarm optimization algorithm based heuristic for the scheduling of workflow applications to cloud resources with an aim to reduce the execution cost by considering the computation cost and data transmission cost. Arabnejad et al. [6] presented a Proportional Deadline Constrained (PDC) for mapping workflow tasks to cloud resources which minimizes the cost while meeting deadline constraints. The algorithm considers the execution cost and time for the selection of resource. Amandeep Verma et al. [7] proposed Budget and Deadline Constrained Heterogeneous Earliest Finish Time (BDHEFT) for workflow scheduling. The spare workflow cost and current task cost are considered for the selection of cost-efficient resource. Abrishami et al. [8] proposed QoS-based workflow scheduling algorithm based on Partial Critical Path (PCP), which tries to minimize the execution cost while meeting user defined deadline. PCP algorithm tries to schedule the critical task that is; the tasks present in the critical path to the resources that executes the task earliest in order to minimize the total cost of the path and executes all the tasks before its finish time. Su et al. [9] proposed a Pareto optimal scheduling heuristic (POSH) to schedule tasks to the cost conscious resource based on pareto dominance. It uses the execution time along with the cost factors to map the higher priority task to the cost efficient resource. Convolbo et al. [10] proposed cost aware scheduling algorithm for solving cost optimization problem for DAG scheduling on IaaS cloud. It schedules the job to the cost efficient resource by computing the execution time and resource usage cost.

3 System Model

3.1 Application Model

A workflow application is modeled by DAG is defined as W = G(T,E), where T is the set of n task$\{t_1,t_2,\ldots.t_n\}$ and E represents the set of directed edges $\{e_1,e_2\ldots.e_k\}$ between the workflow tasks. A task t_i ε T, represents a task in workflow application and each edge ($t_{i,}$ t_j) = e_1 ε E, corresponds to a precedence constraints, where t_j ε T cannot be executed till t_i ε T finishes its execution. Each task t_i ε T represents a computational workload, Wl_i which takes millions of instructions (MI) as a unit of measurement. A task with no predecessor and successor tasks is called as entry task and exit task respectively and the workflow size is determined by the number of tasks [12].

3.2 Cloud Model

Workflowsim [13] is a toolkit used in this experiment to mimic the cloud computing infrastructure. The service provider offers heterogeneous computational resources in the form of virtual machines VM {VM $_1$, VM $_2$,…VM_m} with different prices. Each resource (Virtual Machine) VM_m ε VM is capable of executing the given workflow application and its processing power is expressed in Millions of Instructions per Second (MIPS). Each VM has a different number of cores, MIPS, memory and storage configurations. Pricing is based on pay per use strategy similar to commercial clouds, where the users are charged based on the time interval and type of the resource used.

4 Optimized Cost Scheduling Algorithm

OCSA is an online scheduling algorithm, which comprises of three phases to schedule the biomedical workflow application in the cloud. The phases include Task selection, Resource Selection, and Resource allocation. Task selection phase selects the task with maximum execution time by preserving the parent-child relationship of a given biomedical workflow application. Resource selection phase is a significant phase of this proposed work, as it selects the optimal resource for the task execution and Resource allocation phase allocates the chosen resource to the workflow task for execution. The resource allocation phase in OCSA is a crucial phase where the actual scheduling occurs.

The main objective of the proposed work is to reduce the execution time and monetary cost of Biomedical Workflow applications in a cloud environment. Monetary cost includes execution cost, communication cost, storage cost and resource usage cost [14]. The following time factors are computed before resource selection which in turn is used to compute the various costs resulting in monetary cost for the selection of optimal VM.

$$CT_{t_i} = \frac{L_{t_i}}{VM_c} \tag{1}$$

where CT_{ti} is the Computation time of task t_i which calculates the Computation time of the task by the length of the task L_{ti} with the capacity of Virtual Machine $VM_{c.}$

Data transfer time between the interdependent tasks in a workflow application is calculated as

$$CMT_{t_i} = \frac{\sum_{FS=0}^{t} FS}{VM_b} \tag{2}$$

where $CMT_{ti,}$ represents the communication time of the task t_i, which computes the Communication time between the tasks by the input and output file sizes, FS with Virtual Machine bandwidth, $VM_{b.}$

Expected Execution Time of the workflow task is calculated from the Eqs. (1) and (2), as follows

$$EET_{t_i} = CT_{t_i} + CMT_{t_i} \tag{3}$$

where EET_{ti} is the Expected Execution Time of the task $t_{i,}$ which computes the Expected Execution Time of the task t_i on the VM by the computation time of the task CT_{ti} with the communication time of the task, CMT_{ti}.

Total execution cost for the workflow is calculated by using the execution time with resource usage cost, memory cost, communication cost and storage cost.

$$EC = ET_{t_i} \times C_r \tag{4}$$

where EC is the Execution Cost which computes the Cost for Execution of the task on the VM by the execution time ETt_i with the resource cost C_r.

$$MC = FS \times ET_{t_i} \times C_\mu \tag{5}$$

where MC is the Memory cost which computes the Cost for Memory Usage by the File size FS of the respective task along with its execution time and the memory usage cost, $C_{\mu.}$

$$CC = FS \times CT_{t_i} \times C_\beta \tag{6}$$

where CC is the communication cost which calculates the communication cost between the tasks by the input and output size of the files FS with the time of communication, CTt_i and Bandwidth Cost, $C_{\beta.}$ Communication cost is applicable, only when there exist a dependency between the tasks that is when e_i,e_j > 0. And the communication cost will be zero for the tasks executing on the same resource.

$$SC = FS \times C_s \tag{7}$$

Storage Cost, SC is computed by the size of the file stored with the Storage cost, C_s. And finally, the Minimum Execution Cost is computed from the Eqs. (4)–(7), which is used for selecting the appropriate resource from the heterogeneous resource pool.

$$MEC_w = EC + MC + CC + SC \tag{8}$$

where MEC_w is the Minimum Execution Cost required to execute the Workflow task on the VM.

```
Input: W = G(T,E)
Output: Cost Optimized Workflow schedule
Let n be the number of workflow tasks to be scheduled
    //Task Selection Phase
Task list contains number of tasks $T[n] = \{t_1,t_2,\ldots.t_n\}$
for each task $t_i$ from the tasklist T, Where i = 0 to n
    Select the task with maximum length as $t_{max}$
End for
    //Resource Selection Phase
Resource list contains m Virtual Machines $VM[m] = \{VM_1,VM_2,\ldots.VM_m\}$
for each VM $vm_j$ from the resource list VM, where j = 1 to m
    $VM_{Id}$ <- $VM_j$.getID()
    Compute ComputationTime $CT_{ti}$ of the task $t_{max}$ on $vm_j$ based on Equation (1)
    $CT_{ti}$ <- $L_{ti}/VM_c$
    Compute CommunicationTime ($CMT_{ti}$) of the task $t_{max}$ based on Equation (2)
    $CMT_{ti}$ <- $\sum_{FS=0}^{t} FS$ / $VM_b$
    Compute the ExpectedExecutionTime ($EET_{ti}$) of the task $t_{max}$ on $vm_j$ based on the
    Equation (3)
    $EET_{ti}$ <- $CT_{ti} + CMT_{ti}$
    Compute the Minimum Execution Cost (MEC) required for the execution of task $t_{max}$
    on the $vm_j$ based on the Equation (8)
    MEC <- EC+ MC + CC + SC
    MEcost.add($VM_{Id}$, MEC)
End for
    //Resource Allocation Phase
VMopt <- MEcost.getMinValue()
Set $VM_{opt}$ as BUSY
Allocate $VM_{opt}$ to the task $t_{max}$
```

Algorithm 1. Pseudo code of Optimized Cost Scheduling Algorithm (OCSA)

OCSA selects the appropriate resources for scheduling with the notion to reduce the time and monetary cost of the biomedical workflow applications. It selects the optimal cost VM by considering the execution cost, memory cost, communication cost, storage cost and resource cost so that the workflow tasks are executed in an optimal resource which is shown in the Algorithm 1.

Direct Cost of the applications is measured using the individual resource usage that is data storage cost, resource cost, resource computation cost, Network cost, I/O cost [15, 16]. As the proposed algorithm selects the Optimal VM by computing the direct cost as shown in the Eqs. (4)–(7), it significantly reduces the overall time and cost of execution of workflows in a cloud environment.

5 Experimental Setup and Result Analysis

Workflowsim toolkit is used to create a cloud environment for experimentation purposes, which consists of a Service Provider offering heterogeneous computational resources for workflow execution. The data center configuration is presented in the Table 1, while the resource characteristics and cost for using the resources are shown in Tables 2 and 3 respectively.

Table 1 Datacenter Characteristics

Ram (Host Memory)	20480 MB
Storage (Host Storage)	1000000
Bandwidth	10000
System Architecture	X_86
Operating System	Linux
VMM	Xen

Table 2 Resources characteristics

Resource type	RAM in MB	Processing elements
Small	2048	1
Medium	4096	2
Large	6144	2
XLarge	8192	4

Table 3 Cost of the resources

Resources	Usage cost (Rupees)
Virtual machine	3
Memory	0.05
Storage	0.1
Bandwidth	0.1

For the enhanced analysis and evaluation of the proposed algorithm the experiment is conducted with the real world biomedical workflow applications with a diverse structure and range of tasks varies randomly from 10, 25, 100 and 500. The traces of biomedical workflow applications are downloaded from Pegasus Workflow Generator [11]. The workflow structure of the considered biomedical workflow applications are depicted in the Fig. 1. The proposed algorithm OCSA is compared with the existing workflow scheduling algorithms in workflowsim (FCFS, MAXMIN, MINMIN and MCT) in terms of time and cost.

The execution time of the workflow application is calculated using Eqs. (1) and (2) and the results of OCSA compared with the existing scheduling algorithms are presented in Fig. 2, and the average execution time for the various workflow tasks which results from the average of 20 runs of each workflow execution is compared in Fig. 3,

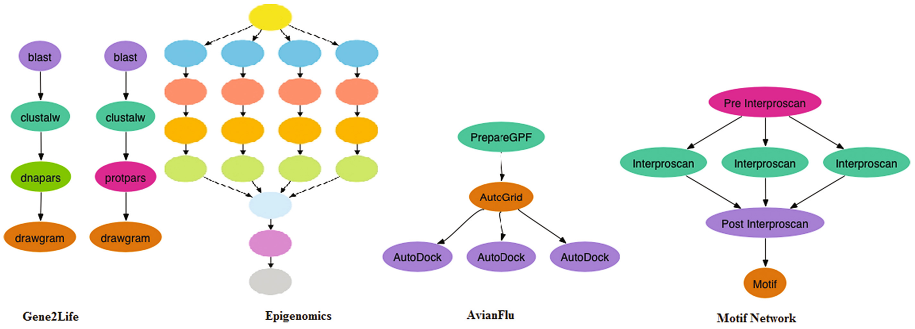


Fig. 1 Structure of the Biomedical Workflow Applications

which substantiate that the execution cost of OCSA is considerably minimum than the other scheduling algorithms.

The execution cost is computed for different biomedical workflow application based on the Eq. (4) and the results are depicted in Fig. 4 and the average cost for different workflow tasks are depicted in Fig. 5, which clearly illustrate that the proposed algorithm performance supersedes the existing approaches and reduces the overall cost.

Optimal Cost Scheduling Algorithm maps the task to appropriate VM by considering monetary costs which are computed by the summation of various individual costs involving storage cost, data transfer cost, memory cost and resource usage cost. This is turn, results in a best scheduling strategy that minimizes the overall execution time and cost of the workflow application.

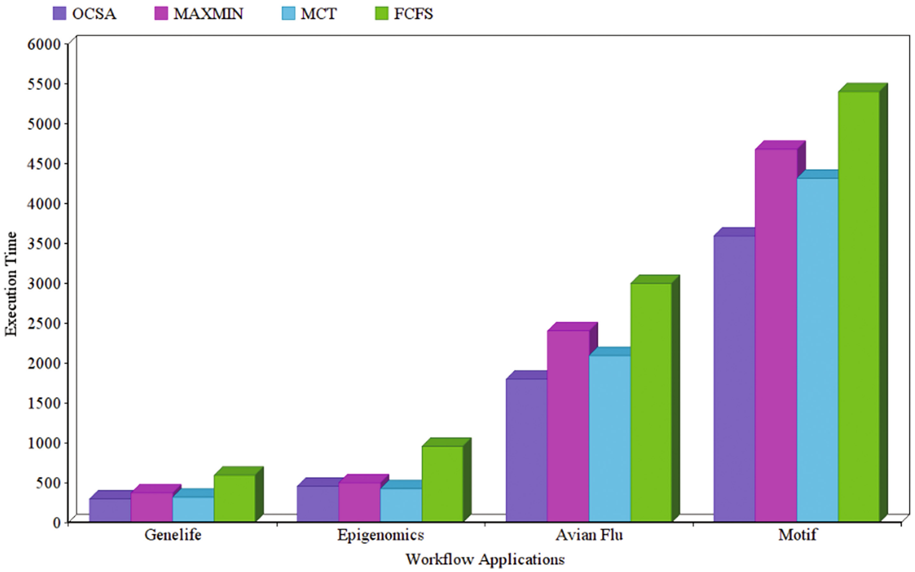


Fig. 2 Execution Time comparisons of biomedical workflow applications

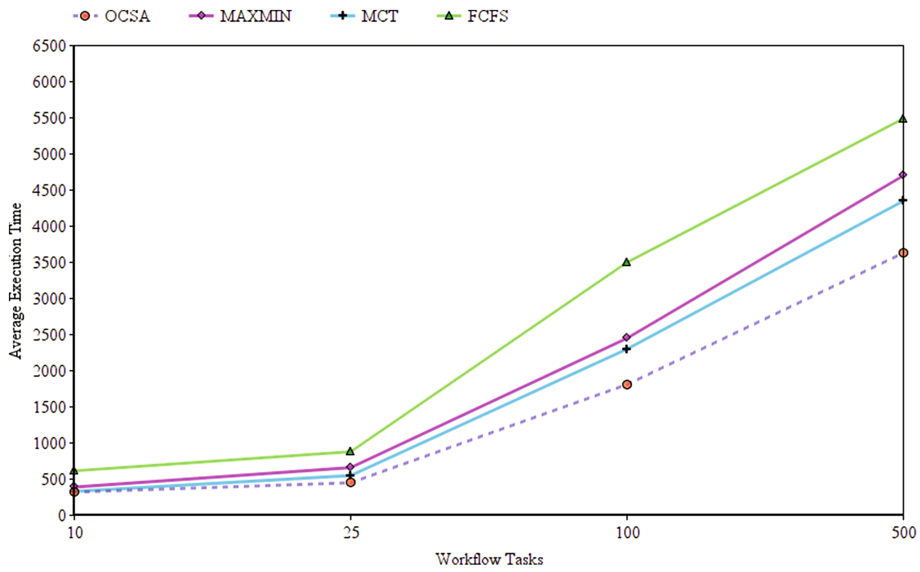


Fig. 3 Average Execution Time comparison of biomedical workflow applications

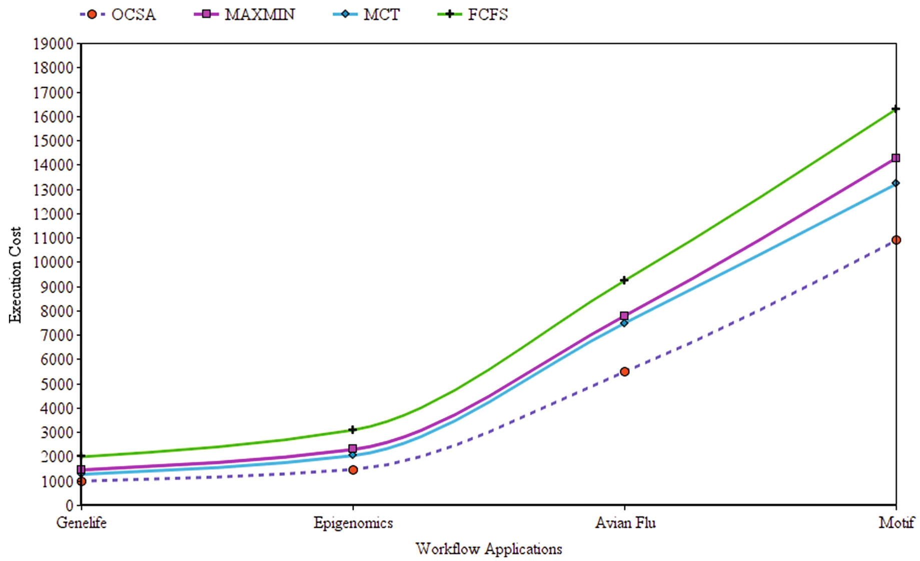


Fig. 4 Execution Cost Comparison of biomedical workflow applications

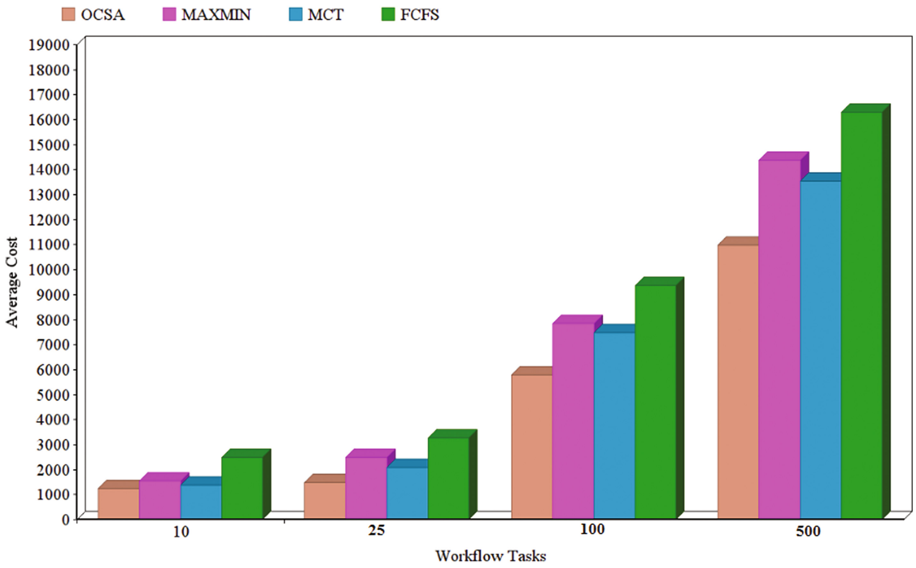


Fig. 5 Average Execution Cost Comparison of biomedical workflow applications

6 Conclusion

An Optimized Cost Scheduling Algorithm is proposed to schedule a biomedical workflow application in cloud with an aim to minimize the overall execution time and cost. The algorithm is evaluated using workflowsim toolkit for four real world biomedical workflow applications and the comparison is made with the existing scheduling approaches of workflowsim (in terms of execution time and cost). The result analysis reveals that the proposed OCSA schedules a workflow application with minimal time and cost in the cloud environment.

References

1. Buyya, R., Pandey, S., Vecchiola, R.: Cloudbus toolkit for market-oriented cloud computing. In: CloudCom 2009 Proceedings of the 1st International Conference on Cloud Computing, vol. 5931. LNCS, pp. 24–44. Springer, Germany, December 2009
2. Armbrust, M., Fox, A., Grifth, R., Joseph, A.D., Katz, R., Konwinski, A., Lee, G., Patterson, D., Rabkin, A., Stoica, I., Zaharia, M.: Above the clouds: a Berkeley view of cloud computing. Technical report, University of California, Berkeley, February 2009
3. Wang, Y., Lu, P.: DDS: A deadlock detection-based scheduling algorithm for work-flow computations in HPC systems with storage constraints. Parallel Comput. **39**(8), 291–305. http://dx.doi.org/10.1016/j.parco.2013.04.006
4. Ullman, J.D.: Np-complete scheduling problems. J. Comput. Syst. Sci. **10**(3), 384–393 (1975)

5. Pandey, S., Wu, L., Guru, S.M., Buyya, R.: A particle swarm optimization-based heuristic for scheduling workflow applications in cloud computing environments. In: 24th IEEE International Conference on Advanced Information Networking and Applications, Perth, WA, pp. 400–407 (2010). doi:10.1109/AINA.2010.31
6. Arabnejad, V. Bubendorfer, K.: Cost effective and deadline constrained scientific workflow scheduling for commercial clouds. In: 2015 IEEE 14th International Symposium on Network Computing and Applications, Cambridge, MA, pp. 106–113 (2015). doi:10.1109/NCA.2015.33
7. Amandeep, V., Sakshi, K.: Cost-Time efficient scheduling plan for executing workflows in the cloud. J. Grid Comput. **13**(4), 495 (2015)
8. Abrishami, S., Naghibzadeh, M.: Deadline-constrained workflow scheduling in software as a service cloud. Sci. Iranica **19**(3), 680–689 (2012). http://dx.doi.org/10.1016/j.scient.2011.11.047
9. Sen, S., Jian, L., Qingjia, H., Xiao, H., Kai, S., Jie, W.: Cost-efficient task scheduling for executing large programs in the cloud. Parallel Comput. **39**(4), 177–188 (2013)
10. Moise, W., Convolbo, J.C.: Cost-aware DAG scheduling algorithms for minimizing execution cost on cloud resources. J. Supercomput. **72**(3), 985–1012 (2016)
11. https://confluence.pegasus.isi.edu/display/pegasus/WorkflowGenerator
12. Mohanapriya, N., Kousalya, G., Balakrishnan, P.: Cloud workflow scheduling algorithms: a survey. Int. J. Adv. Eng. **VII**(III), 188–195 (2016)
13. Weiwei, C., Ewa, D.: WorkflowSim: a toolkit for simulating scientific workflows in distributed environments. In: 8th IEEE International Conference on eScience 2012 (eScience 2012), Chicago, 8–12 October 2012
14. Alkhanak, E.N., Lee, S.P., Rezaei, R., Parizi, R.M.: Cost optimization approaches for scientific workflow scheduling in the cloud and grid computing: a review, classifications, and open issues. J. Syst. Softw. **113**, 1–26 (2016). http://dx.doi.org/10.1016/j.jss.2015.11.023
15. Choudhary, V., Kacker, S., Choudhury, T., Vashisht, V.: An approach to improve task scheduling in a decentralized cloud computing environment. Int. J. Comput. Technol. Appl. **3**(1), 312–316 (2012)
16. Wu, Z., Liu, X., Ni, Z., Yuan, D., Yang, Y.: A market-oriented hierarchical scheduling strategy in cloud workflow systems. J. Supercomput. **63**(1), 256–293 (2013)

Automated Electric Bill Generation System Using Internet of Things

Shreyas H.N., Nikhil Aatrei M, Sumesh S. Iyer, and Prakash P(✉)

M S Ramaiah University of Applied Sciences, Bengaluru 560058, India
hn.shreyas70@gmail.com, nikhilaatrei96@gmail.com,
sumesh1809@gmail.com, prakashp.cs.et@msruas.ac.in

Abstract. Smart City vision attempts to make human life easy. To achieve this, technology has to be used meticulously. Digital service is one of the most vital parts of Smart City. Technology is applied to enhance quality of human life and make it hurdle-less. Human intervention is still a necessity when it comes to electricity bill generation and payment for each household. This research paper proposes an automatic electric bill payment system which reduces human effort. Further the proposed solution supports the smart city vision. The system measures electrical units that is being consumed by the consumers, using Raspberry Pi, and it sends consumption data to the supplier's cloud server wirelessly. The consumption cost will be calculated at the server as per government norms. An e-bill is generated at the server and delivered to the customer via developed Android application. The system has been tested in real time and the usage statistics displayed on the Android application is verified.

Keywords: IoT · Smart city · Raspberry pi · Android · Wi-Fi

1 Introduction

The word smart has been appended with technology, cities. Smart City is all about reducing human effort and improving day to day activities using technology. Migration of people from villages, towns to city has increased drastically [1]. Due to this the infrastructure, requirement of resources has increased across the city. Electricity is one of the vital resources needed.

In the recent era, developments in the field of Internet of Things (IoT) is improving the connectivity between devices of everyday life. IoT can be applied to devices which are equipped with microcontroller, sensors, transceivers, transmitters for digital communication, data sensing and data acquisition. Internet of Things attempts to make the internet more wide-spread and immersive. Given the ease of access to a broad spectrum of devices, IoT can facilitate applications which use huge amount of diverse data to provide new services. This paradigm is applicable in many domains such as medical aid, industrial & home automation, customer service, mobile health care, etc. [2].

The current electricity billing process is that a lineman must go to every household to obtain the meter reading and generate the bill. This process of bill generation is error prone. Consumer can pay the bill manually or via online portals. Even though e-payment is convenient, it has not replaced the manual payment method yet.

S.M. Thampi et al. (eds.), *Advances in Signal Processing and Intelligent Recognition Systems*, Advances in Intelligent Systems and Computing 678,
DOI 10.1007/978-3-319-67934-1_41

Majority of the consumers still opt for manual payment. Manual bill payment is a tedious and time consuming process as it involves commuting to payment center and waiting in queue to pay the bill. There may be a delay in bill payment because of various reasons. For instance, the customer may be compelled to go out of station. In case consumer loses the printed bill, he/she will have to commute to the payment center to get a duplicate copy. Another drawback with the existing process is that the consumer can get to know his/her consumption cost only at the end of each month.

This paper proposes an automated system for electric bill generation and payment. The system can be integrated with any existing electricity meter. Day to day usage statistics and corresponding cost is delivered to the customer through the developed Android application. This enables judicious energy usage.

Further, the rest of this paper is organized as follows: Sect. 2 presents the related work and literature survey. The proposed system architecture is introduced in Sect. 3. The environmental setup along with components is explained in the Sect. 4. The experimental results are discussed in Sect. 5. Finally, Sect. 6 presents the conclusion and future scope of this project.

2 Related Work

Paper [3] discusses about a system which uses a mobile camera to capture the image of the electricity meter reading. Image processing phase undergoes three steps: (1) Preprocessing, which is responsible for cropping the numeric reading area. (2) Segmentation, which outputs individual digits using horizontal and vertical scanning of the cropped numeric area. (3) Recognition of the meter reading by comparing each segment digit with digits' template.

The automatic bill generation [4] using camera, which is fixed in front of the meter, is used to take a snapshot. The captured image is processed to extract meter reading. The generated bill is sent to the customer via SMS using GSM module. [5] explores the development of GSM based electricity bill generator. [6] is about developing an automated bill generating by extracting electrical meter reading using back propagation neural network for optical character recognition (OCR). The back-propagation algorithm is used to extract the meter reading from Digital Meters which will be very helpful for the operators to generate the final electricity bill as per the usage by the customer. This process is very complicated as there are other efficient ways to extract the reading.

[7] uses an Android application to capture meter reading image and then perform OCR. The result of OCR is sent to the developed Web Application which generates the bill. This bill is sent to customer instantly. The poor-quality image due to lighting conditions might cause processing errors.

3 Architecture and Explanation

3.1 Working of Overall System

Every standard electricity meter contains a ‘cal’ LED which blinks each time a fixed amount of energy consumed. The pulse fed to this LED is tapped and given as input to Raspberry pi. Pi is programmed to count the number of pulses. Using this pulse count, Pi calculates the electrical energy units consumed and sends it to electricity supplier’s server which is hosted remotely. The electricity supplier’s server is connected to customer’s bank portal and customer’s android phone.

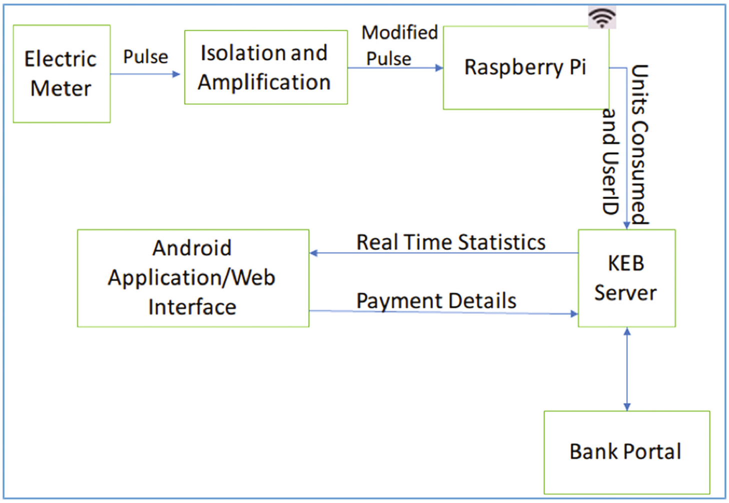


Fig. 1. High level conceptual architecture

The android application which is installed on customer’s cell phone, displays the real-time usage statistics such as day to day consumptions and cost for the current usage. The payment details will be collected by electricity supplier’s server from the customer via android app. The details provided will be authenticated. Later the details are used to connect electricity supplier’s server with customer’s bank portal. The specific bill amount will be deducted from the customer’s account. There are two choices for payment which will be given to customers. The two choices are: 1. Automatic deduction of bill amount on the set date and 2. Manual payment of bill every month via the android application. The automatic setting deducts the amount directly from customer bank account on the given date of every month. In the manual setting, customer can pay within the deadline, manually.

3.2 Internal Working of the System

This section deals with a detailed description of the working of a smart electricity meter.

An electric meter generates a pulse which is an indication of consumption of electricity. This means, 320 pulses will be generated, when 1 kW load is applied for 1 h. An optocoupler is used to electrically isolate the electricity meter from the microcontroller. It is useful because any sort of surge from the meter's side, will not damage the microcontroller and vice versa. A microcontroller (Raspberry Pi 3) is used to count the pulses and upload the count to the server, wirelessly. But since the voltage received from the meter is not strong enough for the microcontroller to detect, that signal is amplified to the required voltage, using an Operational-Amplifier (Op-Amp 741) (Fig. 2).

At the server, the pulse count is received from the microcontroller. The received pulse count is then converted to standard electrical units (KWh) & cost for the consumption is computed. The units consumed and the cost for those many units consumed is stored in a database. This data is used by the server to generate the bill.

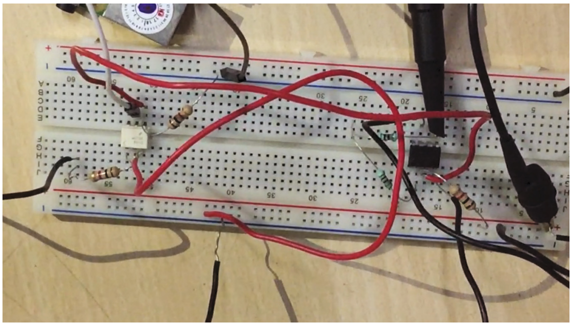


Fig. 2. Circuit containing Op-Amp and Optocoupler

An intuitive android application is developed which is used to get real-time usage statistics & cost corresponding to the usage. The application also facilitates e-payment of the bill.

4 Experimental Environment

This section deals with the components that are required for this experiment and the interconnectivity of these components to match the framework designed in Fig. 1.

4.1 Raspberry Pi

The Raspberry Pi is a minicomputer whose size is similar to that of a computer mouse. It can be used to do a variety of tasks ranging from lighting an LED, from display to image processing. It runs on Linux. It also has additional support of hardware modules like camera. Additional hardware modules could be used as per the requirement of the project. It has inbuilt Wi-Fi capability.

4.2 Optocoupler

It is a device designed to provide electrical isolation between input and output. The isolation of circuit protects from low level, surge and high level noises that could produce errors in the output. Optocouplers and opto-isolators can be used for isolation purpose, or switch to a range of other larger electronic devices such as transistors and triacs providing the required electrical isolation between a lower voltage control signal and the higher voltage or current output signal (Fig. 3)

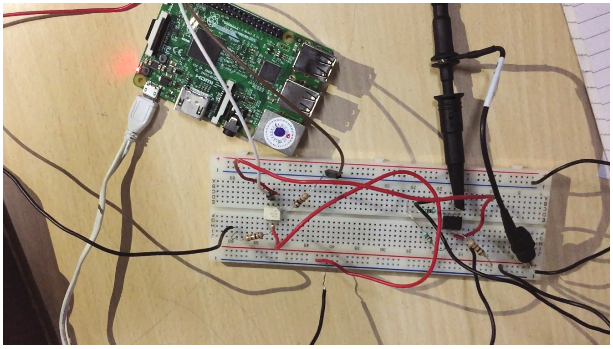


Fig. 3. Interconnectivity of Optocoupler and Raspberry Pi

4.3 Electric Meter

An electric meter is a device that measures the amount of electricity that is consumed by a household. Each meter generates a certain number of pulses per unit consumed. This pulse that the electricity meter generates will be counted by the microcontroller and using this the consumption is measured and cost is calculated based on the consumption (Fig. 4).

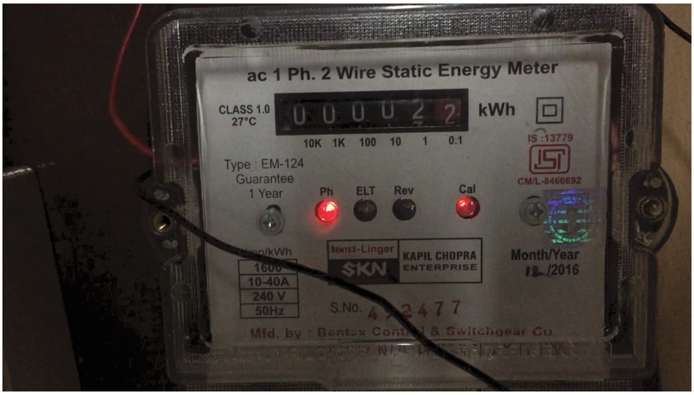


Fig. 4. Electricity Meter with Pulse LED on

5 Experimental Results

5.1 Formulae, Units and Relation Between Important Parameters

Kilo Watt hour is the most commonly used units for measuring electricity consumption. Thus, 1 kWh is considered as ‘1 unit’ of electricity [8]. 1 Unit is thus, the amount of energy consumed in lighting up a 1000-watt bulb for 1 h. The pulse rate of the energy meter being used is 320 pulses/kWh i.e. for every 1000 W of load consumed, there are 320 pulses (blinks) in 1 h.

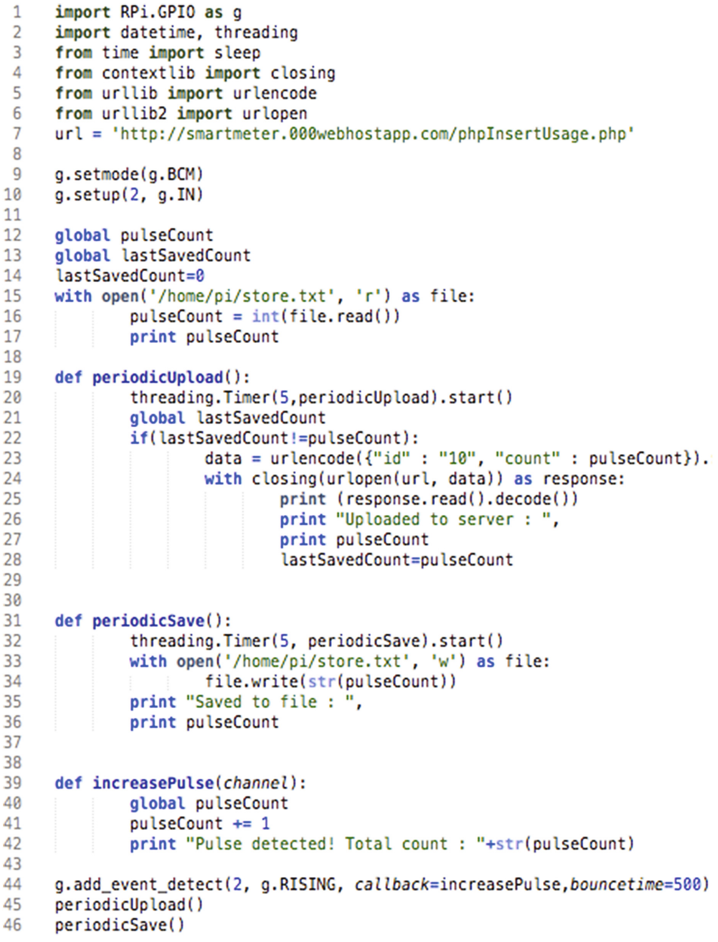
320 pulses → 1000 Wh

1 pulse → 1000/320 Wh/pulse = 3.125 Wh/pulse

So, for 1 pulse 3.125 Watt-Hour is the consumption.

For 1 unit, cost is Rs. 5. Implies that 320 pulses cost Rs. 5.

So, 1 pulse costs 5/320 = Rs. 0.015625/pulse



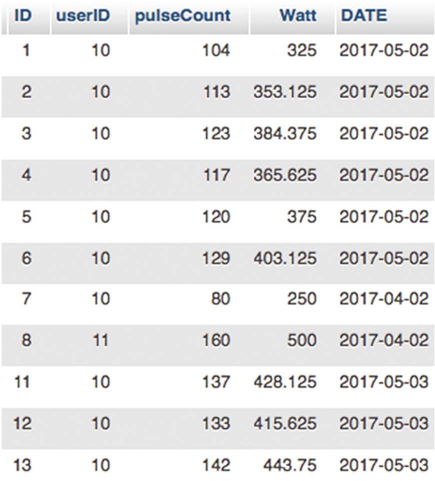
```
import RPi.GPIO as g
import datetime, threading
from time import sleep
from contextlib import closing
from urllib import urlencode
from urllib2 import urlopen
url = 'http://smartmeter.000webhostapp.com/phpInsertUsage.php'

g.setmode(g.BCM)
g.setup(2, g.IN)

global pulseCount
global lastSavedCount
lastSavedCount=0
with open('/home/pi/store.txt', 'r') as file:
        pulseCount = int(file.read())
        print pulseCount

def periodicUpload():
        threading.Timer(5,periodicUpload).start()
        global lastSavedCount
        if(lastSavedCount!=pulseCount):
                data = urlencode({"id" : "10", "count" : pulseCount}).
                with closing(urlopen(url, data)) as response:
                        print (response.read().decode())
                        print "Uploaded to server : ",
                        print pulseCount
                        lastSavedCount=pulseCount


def periodicSave():
        threading.Timer(5, periodicSave).start()
        with open('/home/pi/store.txt', 'w') as file:
                file.write(str(pulseCount))
        print "Saved to file : ",
        print pulseCount


def increasePulse(channel):
        global pulseCount
        pulseCount += 1
        print "Pulse detected! Total count : "+str(pulseCount)

g.add_event_detect(2, g.RISING, callback=increasePulse,bouncetime=500)
periodicUpload()
periodicSave()
```

Fig. 5. Represents code for counting the pulse

5.2 Implementation

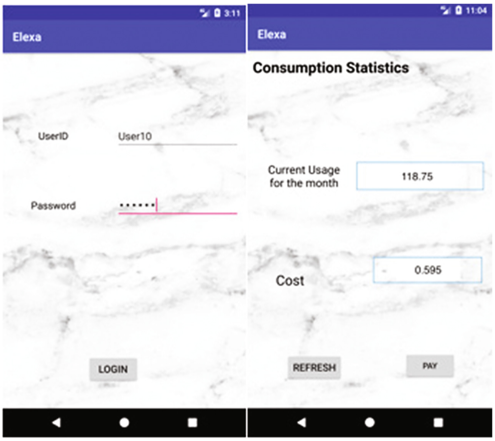
Figure 5 depicts the snippet of code which counts the number of pulses and uploads the counted pulse to a database. The periodicUpload function is responsible for sending data to server with a suitable frequency. The increaseCount function increments count when a rising edge is detected.

Figure 6 depicts the usage details stored at the database. This test data is used for Android application validation. The userID represents unique customer identification number. "Watt" column contains cumulative usage. Real time usage of the customer with userID 10 was tested. The cost was calculated by taking difference between the watt consumed in the late month and latest month end. Referring to Fig. 6, the usage at



ID	userID	pulseCount	Watt	DATE
1	10	104	325	2017-05-02
2	10	113	353.125	2017-05-02
3	10	123	384.375	2017-05-02
4	10	117	365.625	2017-05-02
5	10	120	375	2017-05-02
6	10	129	403.125	2017-05-02
7	10	80	250	2017-04-02
8	11	160	500	2017-04-02
11	10	137	428.125	2017-05-03
12	10	133	415.625	2017-05-03
13	10	142	443.75	2017-05-03

Fig. 6. Test data to check real time statistics obtained at the android application



UserID User10

Password ••••••

Consumption Statistics

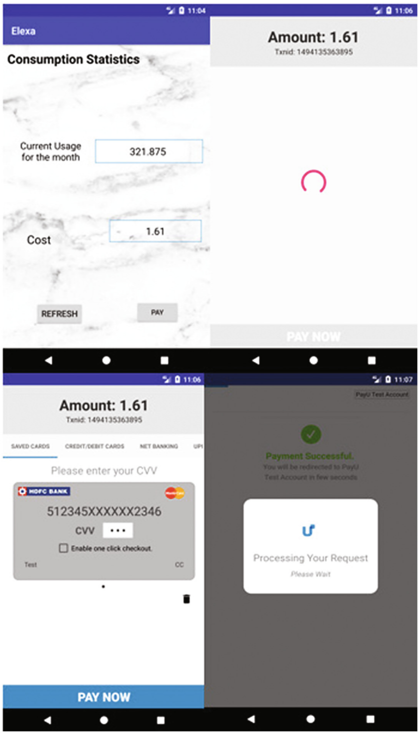
Current Usage for the month 118.75

Cost 0.595

Fig. 7. Login Screen and usage statistics on the Android Application

the beginning of May month = 325 W and the latest usage of May month = 443.75 W. The difference is 443.75–325 = 118.75 W. The corresponding difference in cost 2.22–1.625 = 0.595. The output obtained in the android application is depicted in Fig. 7.

Figure 8 depicts successful payment of the generated bill. This payment details are stored in the database as depicted in Fig. 9.


Elexa

Consumption Statistics

Current Usage for the month: 321.875

Cost: 1.61

Amount: 1.61
Txnid: 1494135363895

Please enter your CVV

512345XXXXXX2346

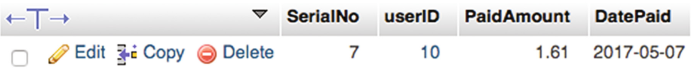
CVV •••

□ Enable one click checkout.

Payment Successful.
You will be redirected to PayU Test Account in few seconds

Processing Your Request
Please Wait

Fig. 8. Payment of Bill using Android Application



←T→	SerialNo	userID	PaidAmount	DatePaid
□ Edit Copy Delete	7	10	1.61	2017-05-07

Fig. 9. Payment Details stored in the Database

6 Conclusion and Future Work

Many solutions are available for the bill automation process using OCR and other technologies but those solutions involve image processing and other methods which are not as accurate. In this paper, the implemented prototype is built on the concept of IoT. The product uses an efficient method to read the consumed electrical energy in terms of pulses and converts them into units as per government norms. Further the system uploads the recorded units to a cloud server. Cost of consumption is then calculated at the server. The bill payment can either be manual or automatic which makes it easy for customers. The customer's bank portal is connected to supplier's server via a secure connection. The prototype developed is well suitable for the concept of Smart City. Analyzing the usage patterns in electricity consumption in a location, can help the electricity distributor to optimize the generation of electricity.

Regarding future work, the device can be improved to support interconnectivity of meters in the locality. The device can be programmed to signal the electricity supplier about unscheduled power outages at a particular area so that immediate action can be taken. The process of complaint registration can be automated. Given that day to day usage details of each consumer is stored at the server, data analysis can be done for various purposes such as optimizing electricity supply. With the help of a Central Smart meter at the distribution transformer, theft of electricity in an area can be detected. The system can be further improved by allowing the supplier to remotely turn off the connection in case of theft or other compelling situations.

References

1. Chandramouli, C.: Rural Urban Distribution of Population. Registrar General & Census Commissioner, India (2017)
2. Bellavista, P., Cardone, G., Corradi, A., Foschini, L.: Convergence of MANET and WSN in IoT urban scenarios. IEEE Sens. J. **13**(10), 3558–3567 (2013)
3. Elrefaei, L.A., Bajaber, A., Natheir, S., AbuSanab, N., Bazi, M.: Automatic electricity meter reading based on image processing. In: Jordan Conference on Applied Electrical Engineering and Computing Technologies. IEEE (2015)
4. Babu, M., Antony, M., Pranav Ashok, N., Niranjana, K.R., Darsana, P.: Automatic electricity billing. Inter. J. Adv. Res. Comput. Commun. **5**(3), 1048–1049 (2016)
5. Jadhav, A.N., Suryavanshi, Y.T., Dewar, B.K., Kumbhar, M.M.: Automatic electric meter reading & monitoring system using GSM. Inter. Res. J. Eng. Technol. **3**(5), 1025–1028 (2016)
6. Shetake, S.R., Patil, A.G.: Automated electricity bill generation by extracting digital meter reading using back propagation neural network for OCR. Inter. J. Eng. Res. Technol. **2**(11), 1348–1350 (2013)
7. Kotwal, J., Pawar, S., Pansare, S., Khopade, M., Mahalunkar, P.: Android App for Meter Reading. Inter. J. Eng. Comput. Sci. (2015). ISSN 2319-7242
8. Bescom. Billing–BESCOM (2017). http://bescom.org/en/frequently-asked-questions-2/billing

Author Index

S.M. Thampi et al. (eds.), *Advances in Signal Processing and Intelligent Recognition Systems*, Advances in Intelligent Systems and Computing 678,
DOI 10.1007/978-3-319-67934-1